Choosing covariates in the analysis of cluster randomised trials

Neil D. Wright

Submitted in partial fulfilment of the requirements
of the Degree of Doctor of Philosophy

Statement of originality

I, Neil Derek Wright, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: 

Date: 23/01/2015

Details of collaboration and publications:

The work reported in Chapter 5 was part of a comprehensive review of the use of covariates in the design, analysis, and reporting of CRTs. This was carried out in collaboration with Noah Ivers, Sandra Eldridge, Monica Taljaard, and Stephen Bremner.

The aims of this review were formed in collaboration with all authors. Noah Ivers and I each independently abstracted information from all papers, and any discrepancies were resolved by discussion. I carried out all data analysis. The interpretation of results was carried out in collaboration with all authors.

A paper from this review is currently in print. The Accepted Author Manuscript is reproduced in Appendix F.

Acknowledgements

Firstly, I wholeheartedly thank Stephen Bremner for his supervision. His guidance, feedback, and patient support have been invaluable. I also thank Sandra Eldridge for her support and insight, which has helped me continually. I could not have wished for better supervisors.

Thank you to the statisticians in the Pragmatic Clinical Trials Unit for their assistance and stimulating discussions, and all staff in the Centre for Primary Care and Public Health for providing a wonderful environment in which to work. Thank you to Richard Hooper and Joan Morris for their useful questions and advice at several points during the project.

I also extend my thanks to Noah Ivers and Monica Taljaard for their collaboration. I found the experience enjoyable and rewarding.

I must express my heartfelt gratitude to my family, especially my parents, Anne and Derek, my grandfather, Ken, and my brother, Paul, for their loving support and faith in me throughout my studies. And thank you to my friends in Essex, the North East, and further afield, for all of your encouragement.

...

Parts of this thesis were proofread for the purposes of spelling and grammar by Alex Raison, Colin Ferguson, and Liam Guest.

This work was supported by the Medical Research Council.

Abstract

Covariate adjustment is common in the analysis of randomised trials, and can increase statistical power without increasing sample size. Published research on covariate adjustment, and guidance for choosing covariates, focusses on trials where individuals are randomised to treatments. In cluster randomised trials (CRTs) clusters of individuals are randomised. Valid analyses of CRTs account for the structure imposed by cluster randomisation. There is limited published research on the effects of covariate adjustment, or guidance for choosing covariates, in analyses of CRTs.

I summarise existing guidance for choosing covariates in individually randomised trials and CRTs, and review the methods used to investigate the effects of covariate adjustment. I review the use of adjusted analyses in published CRTs. I use simulation, analytic methods, and analyses of trial data to investigate the effects of covariate adjustment in mixed models. I use these results to form guidance for choosing covariates in analyses of CRTs.

Guidance to choose covariates a priori and adjust for covariates used to stratify randomisation is also applicable to CRTs. I provide guidance specific to CRTs using linear and logistic mixed models. Cluster size, the intra-cluster correlations (ICCs) of the outcome and covariate, and the strength of the relationship between the outcome and covariate influence the power of adjusted analyses and the precision of treatment effect estimates. An a priori estimate of the product of cluster size and the ICC of the outcome can be used to assist choosing covariates. When this product is close to one, adjusting for a cluster level covariate or a covariate with a negligible ICC provide similar increases in power. For smaller values of this product, adjusting for a cluster level covariate gives minimal increases in power. The use of separate within-cluster and contextual covariate effect parameters may increase power further in some circumstances.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Randomised controlled clinical trials are essential for testing the efficacy and effectiveness of new medical treatments and interventions. Appropriate statistical design and analysis allows valid and useful conclusions to be made. Including data apart from treatment assignment and outcome data in the analysis of a trial may make inference more precise, analysis more efficient, and conclusions more relevant. In a cluster randomised trial pre-existing groups, termed as clusters, of individuals are randomly allocated to treatment arms. This introduces additional complexity to the analysis of the trial and the effects of including other data. This thesis considers the appropriate and effective use of covariate data in the analysis of cluster randomised trials.

1.1 Randomised controlled trials

In a randomised controlled trial (RCT), individuals are randomly allocated to treatment arms. Outcomes are compared between treatment arms to infer the relative efficacy or effectiveness of the treatments. The random assignment to treatment arms ensures there are no systematic differences between individuals in each of the arms at baseline except treatment assignment itself; there are no confounding variables in the classical sense, as described by Hauck et al. [1]. A confounding variable is one that is related to both the outcome and exposure of interest [1]. In an RCT, treatment arm is independent of other variables, and therefore the analysis of RCTs can be based on only the outcome data and treatment arm allocation. For example, we could use a two-sample t-test to compare a normally distributed, continuous outcome variable between two treatment arms and infer a treatment effect. However, in RCTs other variables are often measured and can be used in the analysis of the trial.

1.2 Covariates in randomised controlled trials

A Dictionary of Epidemiology [2] defines a covariate as a "variable that is possibly predictive of the outcome under study", and notes that a covariate "may be of direct interest to the study or may be a confounding variable or effect modifier". In this thesis I consider a covariate to be a variable that is possibly correlated with outcome under study and is independent of treatment arm allocation. Due to randomisation, baseline characteristics measured or fixed (for example, sex and ethnicity) before randomisation fit this definition. This could include demographic characteristics and physiological measurements. Variables measured before randomisation, which may also be measured as outcomes, are also covariates. A variable such as the size of a cluster (the number of individuals in the trial in a cluster) may be a covariate if it is assumed that cluster size is not related to treatment arm (for example, if all individuals are recruited before randomisation).

Many methods for analysing RCTs allow us to include covariate data. In model based analyses of trials this can be done by adding covariate variables and parameters to the linear predictor part of the model. For example, a linear regression model for trial analysis not including covariates could be

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

where Y_i is the outcome variable for individual i, X_i is a binary variable indicating treatment arm allocation for individual i, and the ϵ_i are individual error terms. In this model, the parameter β is the unadjusted treatment effect. A set of covariates Z_1, Z_2, $Z_3 \ldots$ can be included in the analysis model:

$$Y_i = \alpha + \beta X_i + \gamma_1 Z_{i1} + \gamma_2 Z_{i2} + \gamma_3 Z_{i3} + \ldots + \epsilon_i$$

where Z_{ik} is the value of covariate Z_k for individual i. In this model, the parameter β is the treatment effect adjusted for the set of covariates. The adjusted treatment effect parameter represents a treatment effect conditional on values for the covariates.

Adjusting for covariates in the analysis of RCTs can increase the precision of the treatment effect estimate [3] for some models. That is, the uncertainty associated with an estimate of treatment effect may be reduced. Adjusting for covariates can also increase the power of a hypothesis test of no treatment effect: assuming there is a non-zero treatment effect, the probability of a trial producing a statistically significant result is increased [4, 5]. In many non-linear models, such as logistic regression, the unadjusted and adjusted treatment effect parameters are not generally equal [6,7] and have different interpretations [8]. The effects of covariate adjustment in the analysis of RCTs are described in Chapter 2.

Analyses adjusting for covariates are often reported in RCT reports. Reviews indicate between 25% and 72% of trial reports include an adjusted analysis, depending on what trials are in the review and how adjusted analyses are counted [9–11]. There is a range of published guidance on choosing covariates in the analysis of RCTs [12–15], which will be summarised in Section 2.5.

1.3 Cluster randomised trials

A cluster randomised trial (CRT) is an RCT in which pre-existing groups, termed as clusters, of individuals are randomly allocated to treatment arms . This is distinct from individually randomised trials, where individuals themselves are randomly allocated to treatment arms. For example, clusters may be clinical practices or schools where the individuals are patients and school children, respectively.

A cluster randomised design is associated with a loss in statistical power and additional complexity in design and analysis [16]. Therefore, the use of cluster randomisation must be justified by one or more of several reasons [17]. It may be impractical to individually randomise. There may be concerns of contamination between individuals in different treatment arms within the same cluster; for example in the ObaapaVitA trial geographical clusters were randomised, allowing fieldworkers distributing vitamin A supplement or placebo capsules to only carry one type at a time [18]. Cluster randomisation may help compliance with treatment. Routine data may be accessible at cluster level but not at individual level. The treatment may be most appropriately applied at cluster level; for example in a trial of a school-based violence prevention program whole schools were randomised, because all students and staff were necessarily exposed to the intervention [19].

In a CRT there is additional correlation structure in the data, as individuals within the same cluster tend to be more alike than individuals in separate clusters. The intra-cluster correlation coefficient (ICC) quantifies the similarity of a variable for individuals within the same cluster compared to the whole population. The ICC is defined generally as the correlation between values of a variable for individuals in the same cluster [20]. It is assumed that variable values for individuals in different clusters are independent, and that each pair of individuals in the same cluster have the same correlation. That is, the observations within a cluster are exchangeable. It is often assumed that the ICC is the same for all clusters and then the ICC for a variable Y is [20, 21]:

$$\rho = corr(Y_{ij}, Y_{kj}) = \frac{cov(Y_{ij}, Y_{kj})}{\sqrt{var(Y_{ij})var(Y_{kj})}}$$

where Y_{ij} is the value of Y for individual i in cluster j, and $i \neq k$.

It is well established that a valid analysis of a CRT must take account of the clustering, for example by analysing at cluster level, using generalised estimating equations, or using a mixed effects regression model. These methods are widely described, for example by Donner & Klar [22] and Eldridge & Kerry [17].

1.4 Covariates in cluster randomised trials

As with individually randomised trials, in CRTs covariates are independent of treatment arm allocation by design. The analysis of a CRT can be based on only outcome data and treatment arm allocation, plus cluster membership information. A model for the analysis of a CRT with a continuous outcome variable is the linear mixed effects model:

$$\begin{aligned} & Y_{ij} = \alpha + \beta X_j + u_j + e_{ij} \\ \text{where} \quad & u_j \sim N(0, \tau^2) \\ & e_{ij} \sim N(0, \sigma^2) \ . \end{aligned}$$

Y_i is the outcome variable and X_j is a binary variable indicating treatment arm allocation, for individual i in cluster j. The u_j are random cluster effects, and the e_{ij} are independent individual level residuals. The parameter β is the cluster-specific treatment effect. Covariates can be included in this model, just as in a linear regression model, by adding parameters and variables to the linear predictor:

$$Y_{ij} = \alpha + \beta X_j + \gamma_1 Z_{ij1} + \gamma_2 Z_{ij2} + \gamma_3 Z_{ij3} + \ldots + u_j + e_{ij}$$

Z_{ijk} is the value of covariate Z_k for individual i in cluster j. The parameters γ_k are the effects of the covariates, and β is the cluster-specific treatment effect parameter adjusted for the covariates.

As with RCTs, adjusting for covariates in the analysis of CRTs can increase precision of treatment effect estimates, and increase the power of a test of no treatment effect [16,23,24]. Cluster randomisation, however, adds complexity to the effects of covariate adjustment. The variance of the outcome variable can be partitioned into variance at the cluster level and variance at the individual level and both of these can be affected by covariate adjustment. Covariates themselves can exist at two levels: cluster and individual. Cluster level covariates describe something about the cluster and are fixed for individuals in the same cluster. Individual level covariates describe something about the individual, and may be correlated within clusters themselves having a non-zero ICC.

Existing published research on the effects of covariate adjustment in CRTs mostly focuses on the use of linear mixed effects models for continuous outcome data [16, 23–

26]. Published guidance for choosing covariates in the analysis of CRTs [16, 23–25] is limited in scope and application. This literature will be explored in Section 3.3.

1.5 Generalised linear models and mixed effects models

Generalised linear models (GLMs) are regression models that can be used to analyse a number of different types of outcome data including continuous and binary variables. A GLM has three components:

1. The probability distribution, from the exponential family, of the response variable Y.
2. A linear predictor, $\eta = \beta\mathbf{X}$ where $\mathbf{X}$ is a vector of predictor variables and β is a vector of parameters.
3. A link function g such that $E[Y] = \mu = g^{-1}(\eta)$.

Common link functions are linear, log, logit, probit, complementary log-log, and generalised logistic ($g(\mu) = \log(\mu^\theta/(1-\mu^\theta))$) link functions. We are interested in the situation where Y is the outcome of an RCT and there are two predictor variables: X, a binary variable indicating treatment arm; and Z, a covariate independent of X by randomisation. To consider covariate adjustment in GLMs we compare treatment effect parameters and their estimators when using two different linear predictors:

1. $\eta = \alpha^* + \beta^* X$
2. $\eta = \alpha + \beta X + \gamma Z$

Linear predictor (1) gives unadjusted models while linear predictor (2) gives models adjusted for the covariate Z. So β^* is the unadjusted treatment effect parameter, while β is the treatment effect parameter adjusted for Z. The effect of adjusting for a covariate Z on $(\beta^* - \beta)$, the relative precision of treatment effect estimators ($\hat{\beta}^*$ and $\hat{\beta}$), and the power of tests of no treatment effect based on the two models, are summarised in Chapter 2.

A mixed effects model is a statistical model that includes both fixed and random effects. Mixed effects models can be used for the analysis of CRTs by including a random cluster effect. To consider covariate adjustment in mixed effects models we compare treatment effect parameters and their estimators when using the linear predictors:

1. $\eta = \alpha^* + \beta^* X + u_j$
2. $\eta = \alpha + \beta X + \gamma Z + u_j$

The u_j are random cluster effects with $u_j \sim N(0, \tau^2)$. Linear predictor (1) gives unadjusted models while linear predictor (2) gives models adjusted for the covariate Z. β^* is the unadjusted treatment effect parameter, while β is the treatment effect parameter adjusted for Z. Both treatment effects are conditional on the random cluster effect, and so the treatment effects are said to be cluster-specific. Having introduced RCTs, CRTs, covariates, generalised linear models, and mixed effects models, I can now outline the aims of this research project and the structure of this thesis.

1.6 Aims

The overall aim of this project is to address gaps in research on the effects of covariate adjustment in CRTs and produce guidance for choosing covariates in the analysis of CRTs. The project has three major aims:

1. To review existing knowledge and practice in the choice and handling of covariates in the analysis of CRTs.
2. To review published recommendations on choosing covariates, in the context of CRTs and identify which advice is applicable in the analysis of CRTs.
3. To develop further guidance for choosing covariates in the analysis of CRTs.

To address the first aim, I review methodological research on the effects of covariate adjustment in the analysis of individually randomised trials and CRTs. I also summarise published guidance on choosing covariates, and relate this to the methodological research. In the context of that literature, I review current practice in covariate adjustment and choosing covariates in CRTs.

To address the second and third aims, I use data from two published CRTs to investigate distributions of and relationships between covariates and outcome variables. Simulation studies are used to investigate the effects of covariate adjustment in CRTs and evaluate covariate adjustment strategies. The literature reviews, review of current practice, and exploration of trial data are used to develop the plans for these simulations studies. I also present a short analytic investigation of the effects of covariate adjustment in linear mixed effects models.

Results of the literature review and simulation studies are then used to identify which guidance on choosing covariates in individually randomised trials is applicable to

CRTs. They are also used to produce guidance on choosing covariates that is specific to the analysis of CRTs. This guidance is relevant for those designing and analysing CRTs. The work of this project focuses on the analysis of CRTs using linear and logistic mixed effects models.

1.7 Structure of this thesis

In Chapters 2 to 5 I address the first major aim of the thesis. I review the published research on adjusted analyses and choosing covariates in individually randomised trials in Chapter 2, and the equivalent literature for CRTs in Chapter 3. In Chapter 4 I review the two main methods used to investigate the effects of covariate adjustment. In Chapter 5 I review the use of adjusted analyses and choosing covariates in a large sample of published CRT reports.

In Chapters 6 to 10 I address the second and third major aims of the thesis. Chapter 6 contains methods and results of an exploration of data from two published CRTs. In Chapter 7 I outline the methods and plans for simulation studies on the effects of covariate adjustment in CRTs, and the results are given in Chapter 8. In Chapter 9 I present a small analytic investigation of power in adjusted linear mixed effects models. Guidance for choosing covariates in CRTs is discussed in Chapter 10.

In Chapter 11 I discuss the work in this thesis and identify areas for future research.

Chapter 2

Adjusted analyses and choosing covariates in individually randomised trials

I begin with a review of the published literature on adjusted analyses and choosing covariates in individually randomised trials. This provides a useful starting point as there is a large amount of published methodology work which has led to published guidance for choosing covariates. I move on to the literature concerning CRTs in Chapter 3.

The effects of covariate adjustment on treatment effect estimates, the precision of those estimates, and power for generalised linear models are summarised in Section 2.1. The choice of treatment effect estimator, covariates used in randomisation, and baseline imbalance of covariates are considered in Sections 2.2 to 2.4. This literature has led to published guidance on choosing covariates in RCTs, which is summarised in Section 2.5.

2.1 Effects of covariate adjustment in individually randomised trials

The primary analysis of an RCT typically involves estimating a treatment effect, providing a measure of precision for that estimate (as a standard error or confidence interval), and reporting a significance test of no treatment effect. The effects of adjusting for covariates on each of these three results can be considered as follows:

1. The difference between the unadjusted and adjusted treatment effect (Section 2.1.1).

2. The precision of adjusted and unadjusted estimators (Section 2.1.2).

3. The power of tests of no treatment effect based on either the adjusted or unadjusted models (Section 2.1.3).

I focus on analysis using generalised linear models (GLMs) which were introduced in Section 1.5. I am interested in the situation where there is an outcome Y and two predictor variables: X, a binary variable indicating treatment arm; and Z, a covariate. The covariate Z is independent of X by randomisation. To consider covariate adjustment in GLMs we compare treatment effect parameters and their estimators when using two different linear predictors:

1. $\eta = \alpha^* + \beta^* X$

2. $\eta = \alpha + \beta X + \gamma Z$

Linear predictor (1) leads to an unadjusted estimate of treatment effect, while linear predictor (2) will provide an estimate of treatment effect adjusted for the covariate Z. So β^* is the unadjusted treatment effect parameter, while β is the treatment effect parameter adjusted for Z. The effect of adjusting for a covariate Z on treatment effect ($\beta^* - \beta$), the relative precision of treatment effect estimators ($\hat{\beta}^*$ and $\hat{\beta}$), and the power of tests of no treatment effect based on the two models, are summarised in the following sections.

2.1.1 Difference between treatment effects

For GLMs using the linear predictors described above, the unadjusted treatment effect estimator, $\hat{\beta}^*$, is asymptotically unbiased for the adjusted effect, β, if one of the following is true [6]:

1. There is no treatment effect, $\beta = 0$

2. There is no covariate effect, $\gamma = 0$, or the covariate is fixed, $var(Z) = 0$

3. The link function is linear ($g^{-1}(\eta) = a\eta + b$) or a log link ($g^{-1}(\eta) = c\exp(a\eta) + b$)

The first condition establishes that if there is no treatment effect then the adjusted and unadjusted treatment effect parameters coincide. Therefore, a significance test of

no treatment effect (where the null hypothesis is $H_0 : \beta = \beta^* = 0$) is testing the same null hypothesis when using either the unadjusted or adjusted models.

The second condition shows that if the covariate is not at all prognostic then the adjusted and unadjusted treatment effect parameters coincide. So not adjusting for a covariate that is independent of outcome does not bias the treatment effect estimate.

The third condition demonstrates that there is no bias when a linear or log link function is used, but for some commonly used link functions (such as the logit, probit, complementary log-log, generalised logistic, and power link functions) the unadjusted parameter of treatment effect is biased for the adjusted effect parameter [5–7]. The unadjusted parameter is attenuated (closer to the null value) for "any link function based on an inverse cumulative distribution function ... with a log concave density function" [7] such as the logit, probit and complementary log-log link functions. But the unadjusted parameter is biased away from the null for the power family of links ($g(\mu) = (\mu^a - 1)/a$, for $0 < a < 1$) [7].

The magnitude of this bias increases with larger covariate effect (increasing magnitude of γ) or larger variation of the covariate (increasing $var(Z)$) [5, 7]. When γ or $var(Z)$ are close to zero, there is little difference between the unadjusted and adjusted treatment effect parameters [5, 7]. A simulation study by Negassa and Hanley [27] investigated the effects of omitting a single covariate from a logistic regression. With a treatment effect with odds ratio 2.0, omitting a single binary covariate with an effect less than that of the treatment (that is, odds ratio less than 2.0) the percentage change in estimate of treatment effect when omitting the covariate was less than 5%. As the effect of the covariate increased, the relative bias increased from 2.9% to more than 20% (for odds ratios from 2.0 to 10.0). When the omitted covariate was continuous the relative bias was greater than 10% when the effect per unit increase was equal to the treatment effect and the relative bias increased as the covariate effect increased.

The asymptotic bias, as a result of a link function that is not linear or log, described in this section may not be considered a true bias. Rather, the unadjusted and adjusted treatment effects are two distinct treatment effect parameters which are both valid measures of the effect of treatment. The difference in meaning between the adjusted and unadjusted parameters is discussed in Section 2.2.

2.1.2 Precision

The precision of two estimators can be compared by their asymptotic relative precision (ARP). The ARP of an adjusted estimator of treatment effect, $\hat{\beta}$, to the unadjusted

estimator of treatment effect, $\hat{\beta}^*$, is defined by [3]

$$ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{var(\hat{\beta})^{-1}}{var(\hat{\beta}^*)^{-1}} = \frac{var(\hat{\beta}^*)}{var(\hat{\beta})} \ .$$

This value quantifies the increase or decrease in precision obtained by adjusting for the covariate. An ARP greater than one indicates that the adjusted estimator has greater precision.

For linear regression models it can be shown that [3]

$$ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{1 - \rho^2_{XZ}}{1 - \rho^2_{YZ.X}}$$

where ρ^2_{XZ} is the square of the simple correlation of the treatment variable and the covariate and $\rho^2_{YZ.X}$ is the square of the partial correlation of the outcome and the covariate conditional on the treatment. In an RCT $\rho^2_{XZ} = 0$, since patients are randomised to treatment arms, and so

$$ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{1}{1 - \rho^2_{YZ.X}} \geq 1$$

Adjusting for a prognostic covariate (a covariate with $\rho_{YZ.X} \neq 0$) increases the precision of the estimate of treatment effect in a linear regression model, and the stronger the association between the covariate and the outcome, the greater the increase in precision [3].

Robinson et al. [5] consider covariate adjustment in generalised linear models with a binary outcome variable and two binary predictor variables, using seven different link functions. Considering the results which are applicable to the case of randomised trials (when the two predictor variables X and Z are independent), for linear and log link functions $ARP(\hat{\beta} \text{ to } \hat{\beta}^*) \geq 1$. $ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = 1$ if and only if Z is also independent of Y. When using linear or log link functions, the adjusted estimate of treatment effect has greater precision than the unadjusted estimate. However, for models with logit, probit, log-log, complementary log-log or generalised logistic (with $0 < \theta \leq 1$) link functions, $ARP(\hat{\beta} \text{ to } \hat{\beta}^*) \leq 1$. Again $ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = 1$ if and only if Z is also independent of Y. This result is also shown for the logistic regression model by Robinson and Jewell [3]. When using these link functions, the unadjusted estimate of treatment effect has greater precision than the adjusted estimate, in contrast to the result for linear and log link functions.

2.1.3 Power

The power of a significance test of no treatment effect is the probability of a statistically significant result assuming that there is a true treatment effect of a given size.

For a GLM with any link function, $\beta = 0$ implies that $\beta^* = 0$ (Section 2.1.1). So the null hypothesis of a test of no treatment effect can always be stated as $H_0 : \beta = \beta^* = 0$.

As described previously, for linear regression models adjusted and unadjusted treatment effect parameters coincide while the adjusted estimator has greater precision. We conclude that a test of no treatment effect based on the adjusted model has greater power than a test based on the unadjusted model [3].

We can compare tests of no treatment effect by their asymptotic relative efficiency (ARE). Suppose we have a parameter of interest θ and we wish to test $\theta = \theta_0$. And we have a statistic T that is a consistent estimator of $\mu_T(\theta)$, a monotone function (at least close to $\theta = \theta_0$) of θ. A definition of ARE called the Pitman efficiency of two such statistics T_1 and T_2 that are asymptotically normally distributed with means $\mu_{T_1}(\theta)$ and $\mu_{T_2}(\theta)$ and variances $\sigma^2_{T_1}$ and $\sigma^2_{T_2}$ is given by [28]:

$$ARE(T_1 \text{ to } T_2 \text{ at } \theta = \theta_0) = \left[\frac{\mu'_{T_1}(\theta_0)}{\mu'_{T_2}(\theta_0)}\right]^2 \left[\frac{\sigma^2_{T_2}(\theta_0)}{\sigma^2_{T_1}(\theta_0)}\right] \tag{2.1}$$

We can test the null hypothesis $\beta = 0$ using the adjusted or unadjusted Wald test statistics

$$\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})}} \qquad \text{or} \qquad \frac{\hat{\beta}^*}{\sqrt{var(\hat{\beta}^*)}}$$

which have asymptotic normal distributions with means $\beta/\sqrt{var(\hat{\beta})}$ and $\beta^*/\sqrt{var(\hat{\beta}^*)}$ and variances of one. Substituting the adjusted and unadjusted Wald test statistics into the definition of ARE in equation 2.1 we obtain the ARE of a test of no treatment effect based on the adjusted estimator to a test based on the unadjusted estimator [5]:

$$ARE(\hat{\beta} \text{ to } \hat{\beta}^*) = \left[\lim_{\beta \to 0} \left(\frac{d}{d\beta}\beta\right) \left(\frac{d}{d\beta}\beta^*\right)^{-1}\right]^2 \left[\lim_{\beta \to 0} \frac{var(\hat{\beta}^*|X)}{var(\hat{\beta}|X)}\right] \tag{2.2}$$

where we are considering β and β^* as functions of β.

For any link function, $ARE(\hat{\beta} \text{ to } \hat{\beta}^*) \geq 1$ (shown for binary predictor and outcome variables by Robinson et al. [5] and generally by Neuhaus [4]). So adjusting for a covariate independent of treatment allocation increases the power of a test of no treatment effect when using any GLM. The gain in power as a result of adjusting for a prognostic covariate increases with the variance of the covariate effect γZ [4].

A simulation study of logistic regression by Negassa and Hanley [27] estimated power by the proportion of simulated studies in which a hypothesis of no treatment effect was rejected at the 5% level. For the treatment effect, an odds ratio (OR) of 2.0 was used throughout. For a binary covariate, with OR=2.0, power was close to 1 when the covariate was either included or excluded (for sample size 500 or 1000). For a

sample size of 500 and with a covariate OR=10.0 the power was close to 0.9 with covariate included but reduced to 0.8 with the covariate omitted. For a sample size of 1000 the power remained above 0.95 for both covariate included and omitted with covariate OR=10.0. For a continuous covariate the reduction in power was much more pronounced. For a sample size of 500 and with covariate OR=10.0 (per unit increase) the power was close to 0.8 with covariate included but reduced to 0.6 with the covariate omitted. For a sample size of 1000 the power was close to 1 with covariate included but less than 0.9 when the covariate was omitted.

A simulation study by Hernandez et al. [29] investigated the effects of omitting a single binary covariate in a logistic regression. The effect on power was quantified by the percentage reduction in sample size achieved by adjusting for the covariate (so that equal power is obtained by the adjusted and unadjusted tests of no treatment effect). The results of simulations with unadjusted covariate effects of OR=2.0, OR=5.0 and OR=10.0, and unadjusted treatment effects of OR=1.4 and OR=1.7, are summarised in Table 2.1. For a larger covariate effect there is a greater reduction in required sample size. The magnitude of the reduction is smaller when there is a larger treatment effect.

		Unadjusted treatment effect	
		OR=1.4	OR=1.7
Unadjusted covariate effect	OR=2.0	3.3%	2.8%
	OR=5.0	13.8%	13.5%
	OR=30.0	45.5%	43.6%

Table 2.1: Reduction in required sample size for a test of no treatment effect when adjusting for a covariate in a logistic regression. (Figures from Table 3 of Hernandez et al. [29])

2.2 Covariate adjustment and model choice

In many GLMs (such as logistic regression) adjusted and unadjusted treatment effect parameters do not coincide (Section 2.1.1). They are both valid measures of treatment effect, but have different interpretations. The unadjusted treatment effect is a marginal effect, and can be described as population averaged. It is a measure of the effect of treatment on the population of patients in the trial. On the other hand, the adjusted treatment effect is a conditional effect. It is a measure of treatment effect conditional on a given value of the covariate. Where we include more than one covariate, the treatment effect is conditional on all covariates. As more covariates are added

to a model the treatment effect being estimated becomes closer to a subject-specific effect [8].

If the only objective of a trial is a test of no treatment effect, then greater power is obtained by using an adjusted model (Section 2.1.3). If estimation of a treatment effect is the primary aim, then the choice between the unadjusted and adjusted treatment effect depends on what is more appropriate in the context and wider aims of the study. An adjusted estimate of treatment effect (closer to the subject-specific effect) may be more relevant to individual patients in a clinical setting, but the unadjusted model may be more relevant to policy makers interested in effects at population level. This choice is highlighted by Hauck et al. [8]. Austin et al. [10] recommend that there is a need for an "informed debate" about the choice between unadjusted and adjusted treatment effects. Lingsma et al. [30] emphasise the advantage of increasing power by adjusting for covariates, but the difference in interpretation between estimates still remains the overbearing consideration [31].

Hauck et al. [8] and Austin et al. [10] also highlight the difficulties in meta-analysis if only adjusted treatment effects are reported. If the choice of covariates varies between analyses, the same adjusted effect is not being estimated by each analysis so it may be inappropriate or misleading to combine them in a meta-analysis. This argument suggests that presenting an unadjusted estimate of treatment effect, as well as any adjusted estimate, may be useful for future meta-analyses.

2.3 Covariates and trial design

Simple randomisation ensures balance of covariates between treatment arms over all randomisations, but in a single trial there may be chance baseline imbalance. Stratifying randomisation, or using an adaptive randomisation procedure such as minimisation, can be used to restrict the extent of imbalance in one or more chosen covariates. This helps to provide more comparable treatment arms and can improve the efficiency of the analysis of the trial [32].

The use of restricted randomisation has implications for analyses. Standard errors are biased upwards [33] and significance tests have non-nominal size [34] when the effect of a covariate used to balance treatment arms is omitted. Therefore, it is recommended to adjust for covariates used in restricted randomisation in an adjusted analysis. In multi-centre and international trials randomisation may be stratified by centre or country for administrative convenience, and the use of random effects for centre may be useful under some conditions [35]. However, it may be expected that centre or country is not related to outcome and so some authors suggest it is not necessary to

include them as a covariate despite being used in randomisation [12,36].

2.4 Balance of covariates between treatment arms

In an RCT, randomisation ensures that there is no systematic imbalance of covariates between treatment arms. Values of any covariate (measured or unmeasured) are drawn from the same distribution for all treatment arms. The independence between treatment arm and covariates ensures that there cannot be any classical confounding as may be found in observational studies, and allows us to attribute any difference in outcome to a difference in the effects of the experimental and control treatments.

However, in a single RCT there may be imbalance of a covariate between treatment arms. This chance imbalance has been used as justification for the inclusion of covariates that are observed to be imbalanced between treatment arms: for example, in 3 out of 39 adjusted analyses reviewed by Austin et al. [37]. However, this justification is invalid and choosing covariates due to baseline imbalance should be avoided [38–40]. Adjusting for a covariate only when it is significantly imbalanced between treatment arms leads to type I error lower than the nominal level [41].

Consider the example given by Senn [38] of an RCT with continuous outcome Y and a continuous covariate Z. Let n_0 and n_1 be the number of patients in the control and experimental treatment arms, respectively. Assume Z and Y have a bivariate normal distribution with known correlation ρ and standard deviations σ_Z and σ_Y. Let β be the treatment effect. So in the control arm:

$$Z, Y \sim N(\mu_Z, \mu_Y, \sigma_Z^2, \sigma_Y^2, \rho)$$

And in the experimental treatment arm:

$$Z, Y \sim N(\mu_Z, \mu_Y + \beta, \sigma_X^2, \sigma_Y^2, \rho)$$

Denote the means of Z and Y in the control and experimental treatment arms by $\bar{Z}_0$, $\bar{Y}_0$ and $\bar{Z}_1$, $\bar{Y}_1$. Then the standardised difference between treatment arms of the covariate at baseline is

$$\Delta_Z = \frac{\bar{Z}_1 - \bar{Z}_0}{\sigma_Z \sqrt{1/n_0 + 1/n_1}}$$

and the standardised difference between treatment arms of the outcome variable is

$$\Delta_Y = \frac{\bar{Y}_1 - \bar{Y}_0}{\sigma_Y \sqrt{1/n_0 + 1/n_1}} \, .$$

These have a joint normal distribution

$$\Delta_Z, \Delta_Y \sim N(0, \Lambda, 1, 1, \rho)$$

where $\Lambda = \beta/(\sigma_Y\sqrt{1/n_0 + 1/n_1})$. In the analysis of a trial, the value Δ_Z would be used to test baseline balance, while Δ_Y would be used in an unadjusted test of no treatment effect. Under the null hypothesis of no treatment effect ($\beta = 0$), the unconditional distribution of Δ_Y is $N(0, 1)$, however conditional on an observed imbalance of the covariate Z ($\Delta_Z = \delta_Z$) the distribution of Δ_Y is $N(\rho\delta_Z, 1 - \rho^2)$. So, an unadjusted test of nominal size α of no treatment effect will have conditional (on an observed distribution of Z) size

$$\phi(\alpha, \rho, \delta_Z) = \Phi\left[\frac{z_{\alpha/2} - \rho\delta_Z}{(1-\rho)^2}\right] + 1 - \Phi\left[\frac{-z_{\alpha/2} - \rho\delta_Z}{(1-\rho)^2}\right]$$

where $z_{\alpha/2}$ is a critical value of the standard normal distribution, and Φ is the standard normal cumulative distribution function. If covariate Z is prognostic, then $\rho \neq 0$ and typically $\phi(\alpha, \rho, \delta_Z) \neq \alpha$. That is, an unconditional test of no treatment effect will typically have non-nominal size conditional on an observed imbalance of a covariate. For example (given by Senn [38]), let $\alpha = 0.05$, $\rho = 0.73$ and $\delta_Z = 1.43$ as may be seen in a clinical trial. Then the covariate Z is not considered significantly imbalanced between treatment arms (P-value of 0.15), but the conditional size of a test of no treatment effect at the 5% level is in fact 0.09.

The logic of adjusting for only significantly imbalanced covariate suggests that when a covariate is balanced it does not affect significance tests, but the above example shows that this is not the case. In general, the imbalance or balance of a covariate is not an adequate criterion on which to base selection for an adjusted analysis. As Senn [38] notes, unadjusted analyses take into account chance imbalance of covariates and are valid in the absence of knowledge about covariates and their distributions.

Further, selecting covariates on the basis of a significant level of imbalance also fails to recognise some other issues. Only covariates that are observed, reported, and tested could possibly be found to be imbalanced and hence selected; and imbalance neglects the size of the effect of a covariate on the outcome.

Note that it is fair to be cautious of unadjusted treatment effect estimates when a highly prognostic covariate is imbalanced between treatment arms. Altman [42] suggests that it is important to show a baseline comparison table, as general information and to allow readers to assess for themselves the validity of an unadjusted treatment effect estimate, taking into account clinical knowledge and common sense as well as baseline imbalance. This is also included in CONSORT guidelines [43].

2.5 Choosing covariates in the analysis of individually randomised trials

It is not necessary to adjust for any covariates not used in restricted randomisation when estimating the treatment effect in the analysis of an RCT. By design, treatment arm and any covariate (measured or unmeasured) not used to restrict randomisation are independent. So an unadjusted estimate of treatment effect is an unbiased estimate of the marginal treatment effect (which may differ from the conditional/adjusted treatment effect under some models), and a significance test of no treatment effect is of nominal size [38]. It is recommended that the report of a trial analysis includes unadjusted results as well adjusted analyses results [9].

However, there are advantages to adjusted analyses. Firstly, including a covariate will adjust for any chance imbalance of the covariate between treatment arms, and so adjust for any chance bias of the treatment effect estimate [38, 44]. Secondly, in some models, such as linear regression, adjusting for a covariate increases the precision of the treatment effect estimate, although for other models (such as logistic regression) precision decreases when introducing a covariate. Finally, when using any GLM adjusting for a covariate increases the power of a significance test of no treatment effect.

When an adjusted analysis is carried out, the primary motivation for choosing to include a particular covariate is that it is known to be a prognostic factor [12–15, 38]. Gains in precision and power increase as the effect size of an included covariate increases. By incorporating highly prognostic factors as covariates, gains in power are maximised. When the outcome measure of a trial can also be measured at baseline, it is recommended to adjust for baseline by including it as a covariate as baseline measures of outcome are often highly prognostic [12, 45].

It is widely recommended that the selection of prognostic covariates is done *a priori*, on the basis of prior knowledge [13,14]. This may be from previous clinical trials, pilot studies, observational studies, or wider medical knowledge. However, in some cases it may not be possible to select highly prognostic covariates prior to the trial, due to a lack of wider knowledge. In this situation, it is recommended by some authors that *post hoc* selection of covariates is appropriate as long as that method is objective and pre-specified [12]. However, the use of variable selection algorithms may lead to misestimated standard errors and invalid analysis [14]. Choosing a covariate because it is observed to be prognostic of outcome in the particular trial data being analysed is not recommended [13, 14].

Selecting covariates due to an observed baseline imbalance is wholly inappropriate and can lead to invalid analysis [38]. Therefore, it is recommended throughout the literature that this practice is avoided [13, 14, 38].

It is recommended that covariates used to balance randomisation (for example, factors used to stratify randomisation) are adjusted for in an adjusted analysis [13–15]. Failure to adjust for covariates used in randomisation leads to incorrect standard errors and significance tests of non-nominal size [34, 35].

The literature offering guidelines and recommendations for choosing covariates in adjusted analyses of individually randomised trials is almost universally in agreement that covariates should be chosen *a priori* and because they are highly prognostic or used in randomisation. There is some argument that *post hoc* selection of covariates can be justified when there is a lack of previous knowledge, where the method is pre-specified. *Post hoc* selection of covariates due to an observed relationship with outcome, or imbalance between treatment arms should be entirely avoided.

2.6 Conclusion

There is a significant body of published literature on the effects of covariate adjustment in the analysis of individually randomised trials, summarised for GLMs in Section 2.1. This work uses analytic methods, simulation studies, and example trial data. Discussion of other considerations such as interpretation of treatment effect parameters, covariates used in the design of a trial, and imbalance of covariates has been summarised in Sections 2.2 to 2.4.

Published guidance on choosing covariates in individually randomised trials is broadly consistent. These recommendations for the selection of covariates are justified by the statistical literature on the effects of covariate adjustment, parameter interpretation, baseline balance, and trial design.

In CRTs, the effects of covariate adjustment and considerations for adjusted analyses are more complex. The literature relevant to adjusted analyses and choosing covariates in CRTs in particular is looked at in the next chapter.

Chapter 3

Adjusted analyses and choosing covariates in cluster randomised trials: literature review

I now move attention to published literature regarding covariate adjustment in the analysis of CRTs, which are the focus of this thesis. Firstly, I summarise the known effects of covariate adjustment in Section 3.1. In Section 3.2, I look at the proposition of using two covariate effect parameters in adjusted analyses. I summarise published guidance for choosing covariates in CRTs in Section 3.3.

3.1 Effects of covariate adjustment in cluster randomised trials

As with individually randomised trials, we are concerned with the effect of adjusting for a covariate on estimates of treatment effect, precision, and power. I firstly consider change in estimate of treatment effect, and then precision and power for linear mixed effects models. Literature concerning repeated measures and cross-sectional designs with binary outcomes is described in Sections 3.1.3 and 3.1.4.

3.1.1 Change in estimate of treatment effect in linear mixed effects models

Raab & Butcher [23] consider the change in estimate of treatment effect when adjusting for covariates in the analysis of CRTs with linear mixed effects models. Consider a

CRT consisting of J clusters each of m individuals. If no covariate is considered, then the unadjusted linear mixed effects model is

$$Y_{ij} = \alpha^* + \beta^* X_j + u_j + e_{ij} \tag{3.1}$$

where Y_{ij} is the outcome for individual i in cluster j and X_j indicates treatment arm ($X = 1$ for the experimental arm, $X = 0$ for the control arm). The u_j are independent cluster random effects with $u_j \sim N(0, \tau^2)$, and the e_{ij} are individual level residuals with $e_{ij} \sim N(0, \sigma^2)$. The marginal ICC of outcome Y is $\frac{\tau^2}{\tau^2+\sigma^2}$. Now consider introducing individual level covariates $Z_1, Z_2, \ldots$, so the mixed effects model becomes

$$Y_{ij} = \alpha + \beta X_j + \gamma_1 Z_{ij1} + \gamma_2 Z_{ij2} + \ldots + u_j + e_{ij}$$

with Y_{ij} and X_j as before, and Z_{ijk} is the covariate value Z_k for individual i in cluster j. These u_j are independent random cluster effects with $u_j \sim N(0, \tau^2_{y|z})$, and the e_{ij} are individual residuals with $e_{ij} \sim N(0, \sigma^2_{y|z})$. The difference between the unadjusted and adjusted treatment effect estimates is

$$\gamma_1(\bar{z}_{1E} - \bar{z}_{1C}) + \gamma_2(\bar{z}_{2E} - \bar{z}_{2C}) + \ldots$$

where $\bar{z}_{kE}$ and $\bar{z}_{kC}$ are the observed means for covariate Z_k in the experimental (E) and control (C) treatment arms [23]. Note that this is the difference between the unadjusted and adjusted treatment effect estimates in a particular trial, and is not the difference between unadjusted and adjusted treatment effect parameters. When a covariate is at the cluster level, the difference may be large since larger imbalance of the covariate can occur due to a small numbers of clusters [23]. However, restricted randomisation can be used to limit imbalance of such covariates between treatment arms.

3.1.2 Precision and power in linear mixed effects models

Raudenbush [24], Murray [16] and Raab & Butcher [23] consider the analysis of CRTs with continuous outcome measures, analysed with a linear mixed effects model, and give some analytic results for the precision of treatment effect estimates. The Optimal Design software (see Appendix E for reference) allows the user to do power studies in CRTs, and the software documentation [26] and other papers [46] give supporting analytic results.

Firstly, I look at analytic results from Raudenbush [24], Murray [16] and Raab & Butcher [23]. Consider a CRT consisting of J clusters each of m individuals, randomised equally to two treatment arms. If no covariate is considered, then the unadjusted mixed effects model is as in equation 3.1. The marginal ICC of outcome Y is

$\rho = \frac{\tau^2}{\tau^2+\sigma^2}$. The variance of the estimate of the unadjusted treatment effect parameter $\hat{\beta}^*$ is then given by [23, 24]

$$var(\hat{\beta}^*) = \frac{4\Delta}{J} \qquad \text{with} \qquad \Delta = \tau^2 + \frac{\sigma^2}{m} \tag{3.2}$$

which can also be written in terms of the marginal ICC, ρ, as [47]

$$var(\hat{\beta}^*) = 4\frac{(\tau^2+\sigma^2)(1+(m-1)\rho)}{mJ} \ . \tag{3.3}$$

Now consider introducing an individual level covariate Z, so the mixed effects model is

$$Y_{ij} = \alpha + \beta X_j + \gamma Z_{ij} + u_j + e_{ij} \tag{3.4}$$

with Y_{ij} and X_j as before, and Z_{ij} is the covariate value for individual i in cluster j. These u_j are independent random cluster effects with $u_j \sim N(0, \tau^2_{y|z})$, and the e_{ij} are individual residuals with $e_{ij} \sim N(0, \sigma^2_{y|z})$.

If we assume that the regression coefficient of the covariate is known, then the variance of the adjusted treatment effect estimate is [23]

$$\frac{4(\tau^2_{y|z} + \sigma^2_{y|z}/m)}{J} \ .$$

When covariate coefficients are estimated from the data, the variance of the estimate of the adjusted treatment effect parameter $\hat{\beta}$ is [23, 24]

$$var(\hat{\beta}|Z) = \frac{4\Delta_{y|z}}{J}\left[1 + \frac{J(\bar{z}_{..E} - \bar{z}_{..C})^2/4}{\Delta_{y|x}SS_{wz}/\sigma^2_{y|z} + SS_{bz}}\right] \tag{3.5}$$

where

$$\Delta_{y|z} = \tau^2_{y|z} + \frac{\sigma^2_{y|z}}{m}$$

$$SS_{wz} = \sum_{j=1}^{J}(z_{ij} - \bar{z}_{.j})^2$$

$$\text{and} \qquad SS_{bz} = \sum_{j\in E}(\bar{z}_{.j} - \bar{z}_{..E})^2 + \sum_{j\in C}(\bar{z}_{.j} - \bar{z}_{..C})^2 \ .$$

The z_{ij} are values of covariate Z for individual i in cluster j. $\bar{z}_{.j}$ is the mean of covariate Z in cluster j. $\bar{z}_{..E}$ and $\bar{z}_{..C}$ are means of the covariate in the experimental treatment arm (E) and control treatment arm (C). So, SS_{wz} is the pooled, within-cluster sum of squares of the covariate, and SS_{bz} is the pooled, between-cluster (within treatment arm) sum of squares of the covariate.

Comparing equations 3.5 and 3.2 we see that the precision of the estimate of treatment effect increases most with covariate adjustment when $\Delta_{y|z}$ is smaller than Δ, and the means of the covariate are similar in each treatment arm [23,24]. In CRTs with random

effects, the total variance of the outcome is reduced by $(1 - R^2_{yz})$ when adjusting for the covariate Z (where R^2_{yz} is the correlation between Y and Z) [16]:

$$(\tau^2_{y|z} + \sigma^2_{y|z}) = (\tau^2 + \sigma^2)(1 - R^2_{yz})$$

This reduction in variance is not necessarily divided proportionally between the cluster level and individual level components of variance. Raab & Butcher [23] state that when there are "large" clusters it is most important to reduce the cluster level variance to improve the precision of the estimate of treatment effect by covariate adjustment. Imbalance of the covariate between treatment arms will decrease the precision of the treatment effect estimate [23].

3.1.2.1 Optimal Design software

The Optimal Design software (see Appendix E for reference) performs power calculations for several designs of CRTs. For a two level CRT, the software allows the inclusion of a cluster level covariate when the outcome is continuous and a linear mixed effects model is used. The software allows the user to plot a graph of power against one of the following: cluster size (m); number of clusters (J); ICC of outcome (ρ); standardised treatment effect (δ); or the proportion of cluster level variation explained by the covariate (R^2). In each case, the remaining parameters and significance level must be specified by the user.

The software documentation [26] shows how the variance results given above extend to power calculation. Consider a CRT as described above and the unadjusted linear mixed effects model as given in equation 3.1. A test of no treatment effect can be carried out with a nested ANOVA giving an F test statistic, defined as

$$F = \frac{MS_{treatment}}{MS_{cluster}}$$

where $MS_{treatment}$ and $MS_{cluster}$ are the observed mean squares between treatment arm and between cluster (within treatment arm), respectively. Using this test statistic is equivalent to fitting the linear mixed effects model using restricted maximum likelihood estimation and using a Wald test [26]. As the number of clusters J increases, the F statistic converges to the ratio of expected mean squares:

$$\frac{E[MS_{treatment}]}{E[MS_{cluster}]} = \frac{m\tau^2 + \sigma^2 + mJ\beta^{*2}/4}{m\tau^2 + \sigma^2} = 1 + \frac{mJ\beta^{*2}/4}{m\tau^2 + \sigma^2}$$

Under the null hypothesis of no treatment effect ($\beta^* = 0$) this ratio is equal to one. So under the null hypothesis, the F statistic follows a central F distribution with 1 and $J - 2$ degrees of freedom ($F \sim F(1, J - 2)$). Under the alternative hypothesis

($\beta^* \neq 0$) the F statistic follows a non-central F distribution with 1 and $J-2$ degrees of freedom and non-centrality parameter

$$\lambda = \frac{\beta^{*2}}{4(\tau^2 + \sigma^2/m)/J} \ .$$

The non-centrality parameter λ is the ratio of the squared treatment effect (β^{*2}) and the variance of the estimate of treatment effect ($var(\hat{\beta}^*)$). The power of a test of no treatment effect is related to the size of the non-centrality parameter. In order to facilitate power calculation, the treatment effect β^* can be replaced by the standardised effect

$$\delta = \frac{\beta^*}{\sqrt{\tau^2 + \sigma^2}} \ .$$

In terms of the marginal ICC of the outcome ($\rho = \frac{\tau^2}{\tau^2+\sigma^2}$) the non-centrality parameter is

$$\lambda = \frac{J\delta^2}{4(\rho + (1-\rho)/m)}.$$

Given values for J, m, δ, and ρ, the non-centrality parameter can be calculated, and hence the power of a test of no treatment effect can be calculated for a chosen significance level.

Adjusting for a cluster level covariate only, the adjusted linear mixed effects model is as in equation 3.4 with $\sigma^2_{y|z} = \sigma^2$. For a normally distributed covariate Z, the (expected) variance of the treatment effect estimate is [24]

$$var(\hat{\beta}|Z) = \frac{4\Delta_{y|z}}{J}\left[1 + \frac{1}{J-4}\right] \ .$$

The F statistic for a test of no treatment effect is defined as for the unadjusted model:

$$F = \frac{MS_{treatment}}{MS_{cluster}}$$

As the number of clusters increases, the F statistic converges to the ratio of expected values of mean squares:

$$\frac{E[MS_{treatment}]}{E[MS_{clusters}]} = 1 + \lambda_z$$

where

$$\lambda_z = \frac{J\beta^2}{4(\tau^2_{y|z} + \sigma^2/m)}$$

Now under the alternative hypothesis ($\beta \neq 0$) the F statistic follows a non-central F distribution with 1 and $J-3$ degrees of freedom and non-centrality parameter λ_z. Again, the treatment effect β can be replaced by the standardised effect

$$\delta = \frac{\beta}{\sqrt{\tau^2 + \sigma^2}} \ .$$

We express the conditional cluster level variance by $\tau^2_{y|z} = (1-R^2)\tau^2$, so R^2 is the proportion of cluster level variance explained by the covariate. Then the non-centrality

parameter can be given by

$$\lambda_z = \frac{J\delta^2}{4((1-R^2)\rho + (1-\rho)/m)}.$$

Given values for J, m, δ, ρ, and R^2 the non-centrality parameter can be calculated, and hence the power of an adjusted (for a cluster level covariate) test of no treatment effect can be calculated for a chosen significance level.

Raudenbush at al. [48] use the Optimal Design software to investigate the effect on power of adjusting for a cluster level covariate in a linear mixed effects model. The authors note that power in this design depends on the correlation between the cluster level covariate and the true cluster level mean outcome. This is distinct from the correlation between the sample cluster level covariate and the sample cluster level mean outcome.

As an alternative to power of a test, the Optimal Design software also allows calculation of minimum detectable effect sizes (MDES). The MDES is the smallest true effect that will give a statistically significant result at a given power and significance level. Bloom et al. [46] give an expression for the MDES in the analysis of a CRT:

$$MDES \approx M_{J-K}\sqrt{\frac{\rho(1-R_c^2)}{\pi(1-\pi)J} + \frac{(1-\rho)(1-R_s^2)}{\pi(1-\pi)mJ}}$$

where there are J clusters randomised each including m individuals. A proportion π of clusters are randomised to the experimental treatment arm. ρ is the marginal ICC of the outcome, R_c^2 is the proportion of cluster level variance explained by the covariates, and R_s^2 is the proportion of subject level variance (within cluster) that is explained by the included covariates. M_{J-K} is a factor based on the t distribution that accounts for degrees of freedom $(J-K)$. When the degrees of freedom exceeds about 20, M is approximately 2.8 for a two-tailed test with 80% power and a 5% significance level.

3.1.2.2 Power in linear mixed effects models

Konstantopoulos [25] considers the case of adjusting for q cluster level covariates, and r individual level cluster-mean centred covariates. As before consider a CRT consisting of J clusters each of m individuals, randomised equally to two treatment arms. Assume that adjusting for q cluster level covariates reduces cluster level variance by a proportion R_C^2, and adjusting for r individual level cluster-mean centred covariates reduces individual level variance by a proportion R_I^2. Then for a two-tailed t test of the null hypothesis of no treatment effect, the power is given by

$$1 - H\left[t_{(\alpha/2,J-2-q)}, (J-2-q), \lambda_z^{\frac{1}{2}}\right] + H\left[-t_{(\alpha/2,J-2-q)}, (J-2-q), \lambda_z^{\frac{1}{2}}\right]$$

where $t_{\alpha/2,d}$ is the two-tailed critical value of the t distribution at the α level with d degrees of freedom, $H(x, \nu, \lambda_z^{\frac{1}{2}})$ is the cumulative distribution function of the non-central t distribution with ν degrees of freedom and non-centrality parameter $\lambda_z^{\frac{1}{2}}$. The non-centrality parameter is given by

$$\lambda_z^{\frac{1}{2}} = \frac{\sqrt{J}\delta}{2\sqrt{(1-R_C^2)\rho + (1-R_I^2)(1-\rho)/n}}$$

with $\delta = \frac{\beta^*}{\sqrt{\tau^2+\sigma^2}}$ and ρ is the ICC of the outcome (see the Appendix of Konstantopoulos [25]). As noted by Konstantopoulos [25], the power can also be computed using the non-central F distribution with 1 and $(J-2-q)$ degrees of freedom, and the non-centrality parameter λ_z. This is equivalent to the result given in the Optimum Design software documentation [26].

Konstantopoulos [25] uses empirical data from studies in education to further investigate the effects of covariate adjustment in CRTs. From this empirical investigation, Konstantopoulos [25] makes several conclusions:

- For an outcome ICC greater than or equal to 0.1 and with individual level covariates cluster-mean centred, cluster-level covariates typically increase power more than individual level covariates. (Especially when there are 20 or more clusters.)
- When individual level covariates are not cluster mean centred (allowing reduction of individual and cluster level residual variance), they typically increase power more than cluster level covariates.
- For outcome ICC values smaller than 0.1, individual level covariates typically increase power more than cluster level covariates (irrespective of any centring).

3.1.3 Repeated measures designs with continuous outcomes

In a repeated measures CRT we obtain outcome data at two or more time points. In this section we consider the case where there are only two measurements: baseline (t=0) and follow-up (t=1). The baseline measure of outcome may be used to adjust the treatment effect by explicitly modelling time effects. That is, the linear predictor of the model includes an effect for time and time by treatment arm interaction. Baseline measures of outcome are an outcome at time equal to zero. The fixed effect of interaction between time and treatment arm is the treatment effect of interest. I now consider models of this type, in both cross-sectional and cohort designs, when a linear mixed effects model is used for analysis. In a cross-sectional design the subjects

measured at each time point may not be the same individuals, but in the cohort design the same individuals are measured at both time points. I will firstly summarise the work of Murray & Blitstein [47] who consider cross-sectional and cohort designs separately. Then I will consider the paper of Teerenstra et al. [49] who give the effect of including baseline measurement of outcome on power (and sample size) in terms of a single parameter.

Firstly, consider the cross-sectional design and the work of Murray & Blitstein [47]. The adjusted model (which models baseline measurement of outcome as well as follow-up) is given by

$$Y_{ijt} = \alpha + \delta X_j + \zeta t + \beta X_j t + u_j + v_{jt} + e_{ij}$$

where Y_{ijt} is the outcome for individual i in cluster j at time t (where $t = 0$ at baseline and $t = 1$ at follow-up) and X_j indicates treatment arm. The u_j are independent cluster random effects with $u_j \sim N(0, \sigma_c^2)$, the v_{jt} are time by cluster random effects with $v_{jt} \sim N(0, \sigma_{ct}^2)$, and the e_{ij} are individual residuals with $e_{ij} \sim N(0, \sigma_e^2)$. The correlation between Y_{ijt} in the same cluster at different time points is given by

$$\rho_c = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_{ct}^2}$$

which is called the "cluster autocorrelation" [49]. The marginal ICC is given by

$$\rho = \frac{\sigma_c^2 + \sigma_{ct}^2}{\sigma_c^2 + \sigma_{ct}^2 + \sigma_e^2} \ .$$

Murray & Blitstein [47] define the "operative ICC" as the ICC that contributes to the variance of the treatment effect estimate (so it is the adjusted ICC). The operative ICC is given by

$$\rho_{op} = \frac{\sigma_{ct}^2}{\sigma_{ct}^2 + \sigma_e^2} \ .$$

The treatment effect β is estimated by the net difference between time by treatment arm means

$$(\bar{y}_{1E} - \bar{y}_{0E}) - (\bar{y}_{1C} - \bar{y}_{0C})$$

where $\bar{y}_{tE}$ and $\bar{y}_{tC}$ are the means of outcome at time t in the experimental (E) and control (C) treatment arms. The variance of that effect estimate is [16]

$$var(\hat{\beta}) = 8\frac{\sigma_{ct}^2 + \sigma_e^2/m}{J}$$

which can also be expressed in terms of the unadjusted ICC and cluster autocorrelation, or the operative ICC as

$$var(\hat{\beta}) = 8\frac{\sigma^2\rho + m\sigma^2(1-\rho)(1-\rho_c)}{mJ} = 8\frac{(\sigma_{ct}^2 + \sigma_e^2)(1 + (m-1)\rho_{op})}{mJ} \qquad (3.6)$$

where $\sigma^2 = \sigma_c^2 + \sigma_{ct}^2 + \sigma_e^2$. Comparing equations 3.3 and 3.6, the $(\tau^2 + \sigma^2)$ term is replaced by $\sigma_{ct}^2 + \sigma_e^2$ and the unadjusted ICC ρ is replaced by the operative ICC

ρ_{op}. The expression is multiplied by an additional factor of two, which reflects that the treatment effect is estimated by the difference between four means rather than two. The cluster autocorrelation ρ_c is positive, which means that $(\sigma_{ct}^2 + \sigma_e^2) < (\tau^2 + \sigma^2)$ and $\rho_{op} < \rho$. For the adjusted model to give smaller variance of the treatment effect estimate than the unadjusted model these reductions must be large enough to overcome the additional factor of two. This occurs if the cluster autocorrelation ρ_c is large enough.

Murray & Blitstein [47] also consider the cohort repeated measures design as follows. Recall that in a cohort design, the same individuals provide outcome data at two or more time points. The adjusted model (which includes baseline measurement of outcome) is given by

$$Y_{ijt} = \alpha + \delta X_j + \zeta t + \beta X_j t + u_j + s_{ij} + v_{jt} + w_{ijt} + e_{ijt}$$

where Y_{ijt} is the outcome for individual i in cluster j at time t (where $t = 0$ at baseline and $t = 1$ at follow-up) and X_j indicates treatment arm. There are five random effects: the u_j are cluster random effects with $u_j \sim N(0, \sigma_c^2)$; the s_{ij} are random subject effects with $s_{ij} \sim N(0, \sigma_s^2)$; the v_{jt} are time by cluster random effects with $v_{jt} \sim N(0, \sigma_{ct}^2)$; the w_{ijt} are time by subject random effects with $w_{ijt} \sim N(0, \sigma_{st}^2)$; and the e_{ijt} are residuals with $e_{ijt} \sim N(0, \sigma_e^2)$. When there is one measure of each subject at each time point (as usual) the random effects w_{ijt} and e_{ijt} cannot be estimated separately.

The correlation between Y_{ijt} in the same subject and different time points (within subject correlation) is

$$\rho_s = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{st}^2 + \sigma_e^2}$$

which is the "subject autocorrelation". The operative ICC ρ_{op} is then given by

$$\rho_{op} = \frac{\sigma_{ct}^2}{\sigma_{ct}^2 + \sigma_{st}^2 + \sigma_e^2} \ .$$

The variance of the treatment effect estimate is [16]

$$var(\hat{\beta}) = 8\frac{\sigma_{ct}^2 + (\sigma_{st}^2 + \sigma_e^2)/m}{J}$$

which can also be expressed in terms of the operative ICC as

$$var(\hat{\beta}) = 8\frac{(\sigma_{ct}^2 + \sigma_{st}^2 + \sigma_e^2)(1 + (m-1)\rho_{op})}{mJ} \ . \tag{3.7}$$

Comparing equations 3.3 and 3.7, the $(\tau^2 + \sigma^2)$ term is replaced by $(\sigma_{ct}^2 + \sigma_{st}^2 + \sigma_e^2)$ and the unadjusted ICC ρ is replaced by the operative ICC ρ_{op}. As in the cross-sectional design, the expression is multiplied by an additional factor of two which reflects that the treatment effect is estimated by the difference between four means. As the cluster

and subject autocorrelation $(\rho_c + \rho_s)$ are positive, then $(\sigma_{ct}^2 + \sigma_{st}^2 + \sigma_e^2) < (\tau^2 + \sigma^2)$ and $\rho_{op} < \rho$. Again, if the autocorrelations are large enough so that the reductions are large enough to overcome the factor of two, then the adjusted analysis will give smaller treatment effect variance than the unadjusted analysis.

Now turn to the work of Teerenstra et al. [49] who consider repeated measures CRTs with both cohort and cross-sectional design (as well as a mixture of these) together, and express the effect of including baseline measurement in terms of a single parameter. The model (which allows inclusion of baseline measurement of outcome) is given by

$$Y_{ijt} = \alpha + \delta X_j + \zeta t + \beta X_j t + u_j + s_{ij} + v_{jt} + w_{ijt}$$

where Y_{ijt} is the outcome for individual i in cluster j at time t (where $t = 0$ at baseline and $t = 1$ at follow-up)and X_j indicates treatment arm. There are four random effects: the u_j are cluster random effects with $u_j \sim N(0, \sigma_c^2)$; the s_{ij} are random subject effects with $s_{ij} \sim N(0, \sigma_s^2)$; the v_{jt} are time by cluster random effects with $v_{jt} \sim N(0, \sigma_{ct}^2)$; and the w_{ijt} are time by subject random effects with $w_{ijt} \sim N(0, \sigma_{st}^2)$. This model is the same as that used by Murray & Blitstein [47] for the cohort design, except now the w_{ijt} term includes all time by subject variation (i.e. the e_{ijt} have been subsumed into the w_{ijt} term). As before, we have cluster autocorrelation ρ_c and subject autocorrelation ρ_s:

$$\rho_c = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_{ct}^2} \text{ and } \rho_s = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{st}^2}$$

This model is adequate to describe both cohort and cross-sectional designs. In a cross-sectional design, $\rho_s = 0$ and all $s_{ij} = 0$. With $0 < \rho_s < 1$, the model is for the usual cohort design.

The variance of the unadjusted treatment effect estimator (that is, only follow-up outcome is included in the model) is

$$var(\hat{\beta}^*) = \left(\frac{1}{\pi_C} + \frac{1}{\pi_E}\right) \frac{\sigma^2(1 + (m-1)\rho)}{mJ} \tag{3.8}$$

where the proportion of clusters in the experimental and control treatment arms are π_E and π_C, and σ^2 is the total variance over all subjects over all clusters ($\sigma^2 = \sigma_c^2 + \sigma_{ct}^2 + \sigma_s^2 + \sigma_{st}^2$). The variance of a change-from-baseline treatment effect estimator (where the outcome variable used is the difference between outcome measured at follow-up and at baseline) is

$$var(\hat{\beta}^{**}) = 2(1 - r)\left(\frac{1}{\pi_C} + \frac{1}{\pi_E}\right) \frac{\sigma^2(1 + (m-1)\rho)}{mJ}$$

where

$$r = \frac{\sigma_c^2 + \sigma_s^2/m}{\sigma_c^2 + \sigma_{ct}^2 + (\sigma_s^2 + \sigma_{st}^2)/m} = \frac{m\rho}{1 + (m-1)\rho}\rho_c + \frac{1 - \rho}{1 + (m-1)\rho}\rho_s \ .$$

The expected values of the unadjusted ($\hat{\beta}^*$) and change-from-baseline ($\hat{\beta}^{**}$) treatment effect estimators are equal, and so are also equal to any combination $a\hat{\beta}^* + (1-a)\hat{\beta}^{**}$. The combination with smallest variance is when $a = r$, as shown in the appendix of Teerenstra et al. [49].

Including baseline measures of outcome (conducting an analysis of covariance (ANCOVA)) the variance of the treatment effect estimate becomes

$$var(\hat{\beta}) = (1 - r^2)\left(\frac{1}{\pi_C} + \frac{1}{\pi_E}\right)\frac{\sigma^2(1 + (m-1)\rho)}{mJ} \ . \tag{3.9}$$

From equations 3.8 and 3.9, we can see that the ratio of variances for an adjusted analysis to an unadjusted analysis is $(1 - r^2)$. So as r increases, the precision of treatment effect estimate increases, the power increases and the required sample size decreases. Note that r is the correlation between a cluster mean at baseline and at follow-up. It is a weighted average of the cluster autocorrelation and subject autocorrelation, and so will lie between the two and be at least as large as the smaller.

Now consider comparing the cross-sectional and cohort designs. For a completely cross-sectional design, there will be no correlation between subjects at baseline and follow-up conditional on cluster. Therefore the subject autocorrelation will be zero ($\rho_s = 0$). The correlation between cluster means r, now denoted as r_{cross} is then

$$r_{cross} = \frac{m\rho}{1 + (m-1)\rho}\rho_c \leq r \ .$$

The ratio of required sample size is then given by (equation 8 of Teerenstra et al. [49])

$$\begin{aligned} \frac{var(\hat{\beta}_{cross})}{var(\hat{\beta}_{cohort})} &= \frac{1 - r^2_{cross}}{1 - r^2} \\ &= \left(1 + (1-\rho)\frac{\rho_s}{(1 + (n-1)\rho)(1-r)}\right) \\ &\quad \times \left(1 - (1-\rho)\frac{\rho_s}{(1 + (n-1)\rho)(1+r)}\right) \end{aligned} \tag{3.10}$$

So the cross-sectional design gives larger variance of treatment effect estimator (or a larger required sample size for the same variance) than an equivalent cohort design, although for a large cluster size or small subject autocorrelation the difference is small.

3.1.4 Cross-sectional designs with binary outcomes

Nixon & Thompson [50] investigate the effect on the precision of treatment effect estimates when adjusting for baseline in the analysis of repeated cross-sectional CRTs with binary outcomes. In such a trial there is a binary outcome measured at baseline and at follow-up, but outcomes are not necessarily measured on the same individuals

so analyses cannot be adjusted for individual baseline values. Instead, baseline proportions for each cluster can be adjusted for in the analysis. Nixon & Thompson [50] use simulations to compare the width of confidence intervals for treatment effect estimates between unadjusted and adjusted mixed effects models. The number of clusters ranged from 10 to 100, and fixed cluster size ranged from 10 to 1000.

When the baseline probabilities were prognostic and there was mild heterogeneity between clusters (a standard deviation of logit probabilities between clusters of 0.1 at baseline), simulations showed baseline adjustment making "little difference" [50] to the estimation of β; mean confidence interval widths were similar between adjusted and unadjusted models [50]. With extreme heterogeneity of clusters (a standard deviation of logit probabilities between clusters of 0.5) at baseline or both baseline and follow-up, then adjusting for baseline proportions improved precision in the estimation of the treatment effect parameter. Nixon & Thompsons [50] repeated simulation studies with variable cluster size. The number of clusters was fixed at 30 while the cluster size was determined by a random variable uniformly distributed between 49.5 and 150.5 and rounded to the nearest integer, to give a range of values around 100 as used in the fixed cluster size case. These analyses gave similar results to when cluster size was fixed at 100 [50]. Simulations where baseline is not prognostic of outcome showed that adjusting for baseline gave less precise estimates when there is a small number of clusters [50]. To consider the performance of adjusting for baseline proportions when there is imbalance at baseline, Nixon & Thompson [50] also simulated data sets where the observed mean of cluster log odds at baseline in the intervention arm is at least $\sqrt{2}$ standard errors above the observed mean of cluster log odds at baseline in the control arm. In this case, baseline adjustment was successful in correcting the chance bias in the treatment effect estimate [50].

Nixon & Thompson [50] conclude that precision is only markedly increased when adjusting for baseline proportions in a repeated cross-sectional CRT when there is a large cluster size and baseline heterogeneity between clusters. The number of clusters affects the overall precision but does not appear to influence the effect of baseline adjustment. Baseline adjustment successfully adjusts for an observed baseline imbalance regardless of cluster size or number of clusters.

Austin [51] investigates the power of methods of adjusting for baseline in the analysis of repeated cross-sectional CRTs with binary outcomes. Austin [51] applies a total of nine models to simulated trial data. There are three analysis methods used, which are GEEs, mixed effects models and cluster level linear regression. For each of these, three different models are used: an unadjusted model, a model adjusted for baseline, and a model estimating change in response log odds or probability.

When the cluster heterogeneity was small the mixed effects model estimating change in log odds gave consistently lower power for a test of no treatment effect than the other models. The differences in power increased as the number of clusters or number of individuals in each cluster increased. Differences in power between the other models were "negligible" [51]. Any minor difference in power decreased as the number of clusters or number of individuals in each cluster increased. When the cluster heterogeneity was large similar results were seen, except the three unadjusted methods had "marginally" [51] less power when the number of individuals in each cluster was large. The differences in power between models was clearer when the cluster autocorrelation was large. All of these results were essentially identical whether there was a zero or non-zero secular time effect.

Austin [51] concludes that with exception of mixed effects model estimating change in log odds, all methods considered had comparable power. However, when the number of individuals in each cluster is large, methods that did not adjust for baseline had less power.

3.2 Within-cluster and contextual covariate effect parameters

Individual level covariates generally have a different value for each individual within a cluster [52]. An individual level covariate, Z_{ij}, can be separated into two components: a between-cluster component $\bar{Z}_j$ (the mean of the covariate in cluster j), and a within-cluster component $(Z_{ij} - \bar{Z}_j)$. The literature considered so far describes the case of using a single variable and single parameter for the effect of a covariate on outcome. This assumes that effects of the covariates at the individual and cluster level are identical, which may not be true [52]. Using a single parameter for a variable that has different effects at the individual level and cluster level leads to misleading estimates for the effect of the covariate, and biased estimates of residual variance [53].

The mixed effects models used for the analysis of CRTs can easily be adapted to incorporate separate individual and cluster level covariate effects. For example, Raudenbush [24] gives the following modified version of equation 3.4:

$$\begin{aligned} Y_{ij} &= \alpha + \beta X_j + \gamma_W Z_{ij} + \gamma_B \bar{Z}_j + u_j + e_{ij} \\ &= \alpha + \beta X_j + \gamma_W \bar{Z}_j + \gamma_W (Z_{ij} - \bar{Z}_j) + \gamma_B \bar{Z}_j + u_j + e_{ij} \\ &= \alpha + \beta X_j + \gamma_W (Z_{ij} - \bar{Z}_j) + (\gamma_W + \gamma_B) \bar{Z}_j + u_j + e_{ij} \end{aligned}$$

Recall, Y_{ij} is the outcome for patient i in cluster j, X_j indicates treatment arm for cluster j, and u_j and e_{ij} are random effects ($u_j \sim N(0, \tau^2)$, $e_{ij} \sim N(0, \sigma^2)$). Instead of

a single covariate effect γ, there are two covariate effects. I will follow the naming convention of Korendijk et al. [54], so γ_W is the *within-cluster* coefficient of the covariate effect and γ_B is the *between-cluster effect*. The *contextual effect* is $\gamma_C = (\gamma_W + \gamma_B)$, and reflects the regression of cluster mean outcome $\bar{Y}_j$ on cluster mean covariate $\bar{Z}_j$. When the between-cluster effect is zero, the within-cluster and contextual effects are equal, which is the assumption of the model with a single covariate effect parameter.

Two simulation studies [54,55] have been carried out to investigate the effects of using separate within-cluster and contextual covariate effect parameters in the analysis of CRTs with linear mixed effects models.

3.2.1 Bias and precision in linear mixed effects models

Korendijk et al. [54] investigate the effects of ignoring distinct within-cluster and contextual effects on parameter estimates and their standard error estimates, in the analysis on CRTs. Korendijk et al. [54] compare a model that includes a within-cluster covariate effect and a between-cluster covariate effect (the "different-effects" model)

$$Y_{ij} = \alpha + \beta X_j + \gamma_W Z_{ij} + \gamma_B \bar{Z}_j + u_j + e_{ij}$$

with a model that includes only a single covariate effect (the "ordinary" model)

$$Y_{ij} = \alpha + \beta X_j + \gamma Z_{ij} + u_j + e_{ij} \ .$$

In each case, u_j are independent cluster random effects with $u_j \sim N(0, \tau^2)$, and the e_{ij} are individual residuals with $e_{ij} \sim N(0, \sigma^2)$, that are different to the u_j and e_{ij} in the previous model.

Cluster size was fixed at 5, with 20 clusters. The ICC of the outcome was fixed at 0.1. The "inequality of covariate effects" (ICE) [54] was defined by the ratio of the within-cluster effect to the contextual effect:

$$ICE = \frac{\gamma_W}{\gamma_W + \gamma_B}$$

So, with $ICE = 1$ the between-cluster covariate effect is zero, and the model including a single covariate effect is correct in assuming a single covariate effect.

Under both analysis models, the treatment effect parameter estimates were unbiased for all values of the "inequality of covariate effects" and total covariate effect. For an "inequality of covariate effects" of 0.1, 27.9% of estimated standard errors from the "ordinary" model were inflated (that is, greater than the 97.5th percentile of the estimated standard errors from the "different-effects" model). For "inequality of covariate effects" of 0.333, 3, and 10, the proportion of inflated standard errors

from model B were 5.7%, 3.1%, and 6.3% respectively. So there was only serious inflation of estimated standard errors under the "ordinary" model when the within-cluster covariate effect was particularly small. However, under the "ordinary" model the coverage of 95% confidence intervals is close to 0.95 under all conditions.

Korendijk et al. [54] conclude that in general, treatment effect estimates and estimated standard errors are estimated without bias in both models. So it is sufficient to include a single covariate effect in the analysis model in this regard. However, Korendijk et al. [54] warn that the parameters associated with random effects (the residual cluster level and individual level variances) are estimated with bias under the model that includes only a single covariate effect.

3.2.2 Power in linear mixed effects models

Klar & Darlington [55] consider the analysis of CRTs where the outcome measured at baseline can be used as a covariate. They assume a cohort repeated measures design, so the baseline outcomes are from the same individuals as follow-up outcome.

Klar & Darlington [55] consider analysis with the following five linear mixed effects models:

1. Model 1 including only follow-up outcome data:

$$Y_{ij} = \alpha + \beta X_j + u_j + e_{ij}$$

2. Model 2 using the simplest method of adjusting for baseline, modelling change in outcome:

$$Y_{ij} - Z_{ij} = \alpha + \beta X_j + u_j + e_{ij}$$

3. Model 3 adjusting for baseline measurement of outcome by including a single covariate effect:

$$Y_{ij} = \alpha + \beta X_j + \gamma(Z_{ij} - \bar{Z}) + u_j + e_{ij}$$

4. Model 4 allowing distinct within-cluster and contextual effects of baseline outcome on follow-up:

$$Y_{ij} = \alpha + \beta X_j + \gamma_W(Z_{ij} - \bar{Z}_j) + \gamma_C(\bar{Z}_j - \bar{Z}) + u_j + e_{ij}$$

5. Model 5 modelling change in outcome and allowing distinct within-cluster and contextual effects of baseline outcome on follow-up:

$$Y_{ij} - Z_{ij} = \alpha + \beta X_j + \gamma_W(Z_{ij} - \bar{Z}_j) + \gamma_C(\bar{Z}_j - \bar{Z}) + u_j + e_{ij}$$

Note that the variances of the u_j and e_{ij} are not the same between models. Where there is a single covariate effect, γ is the effect of baseline outcome on follow-up. In the models allowing distinct within-cluster and contextual covariate effects, γ_W is the within-cluster and γ_C is the contextual effect of baseline on follow-up.

Klar & Darlington [55] simulated data for CRTs containing 28 clusters with 58 individuals in each cluster. ICC for both baseline and follow-up measurements of outcome was 0.03.

Klar & Darlington [55] find that in all models type I error was at the correct level. Model misspecification does not affect type I error. The power when using Model 4 was "consistently high" [55] for all parameter values compared to other models. Model 5 also gave "consistently high" [55] power and identical results to Model 4 when the cluster size was fixed. When there was no effect of baseline measurement of outcome on follow-up ($\gamma_W = \gamma_C = 0$), Model 1 gave higher power (similar to models 3, 4 and 5) while Model 2 gave lower power. When there was a single covariate effect ($\gamma_W = \gamma_C$), models 3, 4 and 5 gave high power. For distinct within-cluster and contextual effects of baseline on follow-up, Model 4 gave higher power than the other models. The relative performance of the other models then depended on the combination of parameter values.

From these results, Klar & Darlington [55] recommend that in CRTs where there is a baseline measurement of outcome for individuals, Model 4 should be used in analysis to adjust for baseline. This model allows distinct within-cluster and contextual effects of baseline outcome on follow-up. Further, use of this model gives estimates of the within-cluster and contextual effects, which may be of interest to the researchers.

3.3 Choosing covariates in the analysis of CRTs

The selection of covariates in the analysis of CRTs is more complicated because covariates exist at both the cluster level and individual level. Further, adjustment for an individual level covariate can affect the residual variance of the outcome at both the cluster and individual levels. The choice of actual covariates will depend of the nature of the trial, but Murray [16] recommends identifying a set of potential covariates that are related to the outcome measure. Then in a preliminary analysis, examine which covariates account for individual variation and include these in the analysis. This can help increase the precision of the estimate of treatment effect [16]. However, Raab & Butcher [23] propose that the arguments used for individually randomised trials for specifying covariates *a priori* are "equally applicable" in the analysis of CRTs.

Raab & Butcher [23] advise that in the best interests of precision and change in estimate of treatment effect, trial analysts should identify a set of covariates to include in analysis that are highly prognostic at the cluster level. Other papers also suggest that reducing cluster level variation by including cluster level covariates is most useful [48, 56]. However, adjusting for a cluster level covariate that is not correlated with outcome compromises power due to the loss of a degree of freedom [25,48], though this is negligible except when sample size is small [48]. Konstantopoulos [25] recommends that cluster level covariates should be carefully chosen to balance benefits in precision with the cost of losing degrees of freedom [25]. Further, Konstantopoulos [25] argues that both individual and cluster level covariates can be effectively used to improve power, depending on assumptions, trial design, and variable distributions.

In linear models, adjusting for a baseline measure of outcome (possibly by explicitly modelling time) can increase precision and power [47]. Murray [16] states that a baseline measurement of the outcome is the "single best covariate" to include. However, from simulation studies of repeated measures CRTs with binary outcomes there appear to be limited gains in precision or power for including baseline measures of outcome. Only when there are large cluster sizes and heterogeneity between clusters at baseline is there an advantage in including baseline [50, 51].

Neuhaus & Kalbfleisch [52] give examples where individual level and cluster level effects of a covariate on outcome are distinct. It is possible to separately model these cluster and individual level effects of covariates when adjusting. Using a single covariate effect when cluster and individual effects differ does not bias the treatment effect estimate or its precision [54]. Using separate covariate effects is recommended by Korendijk et al. [54] to ensure variance parameter estimates are unbiased. Klar & Darlington [55] recommend using separate covariate effects, as this can give greater power for a test of no treatment effect.

Moerbeek [57] compares strategies of recruiting more clusters or measuring, and adjusting for, covariates to increase power. When adjusting for a covariate, separate within-cluster and contextual covariate parameters are used. Moerbeek [57] recommends that when cluster size is small, and the cost to recruit clusters is large, individual level covariates should be measured. When cluster level size is large, but cost to recruit a cluster is small, measuring cluster level covariates is recommended.

3.4 Conclusion

Published research on the effects of covariate adjustment in CRTs is limited, focussing on linear mixed effects models and a limited number of other designs and analysis

methods. Notably, there is little work on the effects of covariate adjustment when using logistic mixed effects models. A small number of authors have considered including separate within-cluster and contextual covariate effect parameters in linear mixed effects models. They suggest that in some cases such a model may be preferred to a model with a single covariate effect parameter. The simulation studies consider only specific sample sizes and values for ICCs, and the same ICC for a covariate and the outcome variable.

Existing recommendations for choosing covariates in the analysis of CRTs are limited in quantity and scope. There is some direct extension of recommendations from individually randomised trials, such as the general advice of choosing highly prognostic covariates *a priori*. However, the issue of choosing covariates in CRTs has rarely been addressed directly and with consideration for practical usefulness of results.

There exists significant scope for further work on the effects of covariates in CRTs, and the development of recommendations for choosing covariates in the analysis of CRTs. In particular, the effects of covariate adjustment when using logistic mixed effects models warrant investigation. There is also potential further development of results outlined in this chapter and Chapter 2 into practical recommendations for researchers undertaking CRTs. The research presented in the later chapters of this thesis addresses these areas for further investigation. That work is preceded, in the next chapter, by a review of the analytic and simulation methods that have been used in previous work to investigate the effects of covariate adjustment in RCTs and CRTs.

Chapter 4

Methods used to investigate the effects of covariate adjustment

In this chapter I review the methods used to investigate the effects of covariate adjustment in individually randomised and cluster randomised trials in the published research that was considered in Chapters 2 and 3. The methods can be separated into two groups: analytic methods, and simulations. Analytic methods involve the use of mathematical analysis, applying known mathematical and statistical theorems, to investigate the theoretical effects of covariate adjustment. Simulation, specifically Monte Carlo simulation, is the generation of large numbers of samples to approximate sampling distributions under a chosen model. I look at the use of each of these approaches, and consider how they may be useful in further research.

4.1 Analytic methods

These methods seek to compare parameters and their estimators between an adjusted model that incorporates a single covariate (independent of treatment arm) and an unadjusted model that includes only a treatment effect. As described in Chapter 2, three separate effects of covariate adjustment are of interest:

1. The difference between the unadjusted and adjusted treatment effect - asymptotic bias.
2. The precision of adjusted and unadjusted estimators.
3. The power of tests of no treatment effect based on either the adjusted or unadjusted models.

Published research has tended to apply analytic methods to each of these separately. This is expected since the asymptotic bias is a feature of only the parameters defined by the model. Precision, however, is a property of estimators of those parameters and power depends on asymptotic properties of both parameters and their estimators.

4.1.1 Difference between treatment effects

Two papers [6, 7] present methods of investigating the effect of omitting a covariate on treatment effect parameters when using GLMs. Gail et al. [6] identify conditions where the difference between unadjusted and adjusted treatment effect parameters is zero and give an approximation for this difference when the covariate effect is close to zero. Neuhaus & Jewell [7] identify conditions where the asymptotic bias is zero, but also find conditions that determine the direction of the asymptotic bias. Further, Neuhaus & Jewell [7] give an expression for the unadjusted treatment effect parameter in terms of the adjusted treatment effect parameter for a treatment effect parameter close to zero.

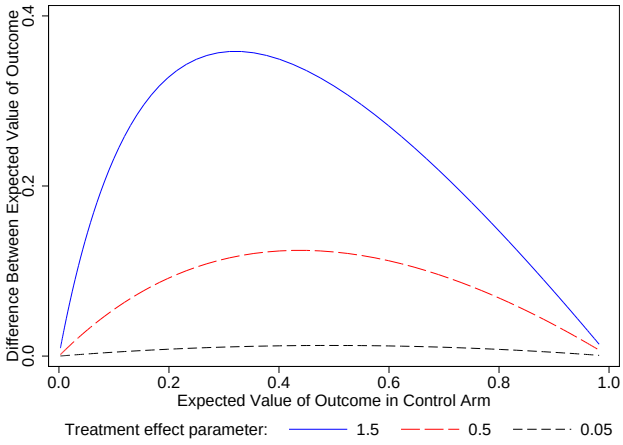
Gail et al. [6] firstly define unadjusted and adjusted models using the conditional expected value of the outcome variable. The authors consider method of moments estimators for the model parameters. It follows from these that the treatment effect parameters coincide if the treatment effect or covariate effect are zero. Gail et al. [6] then use a second-order Taylor expansion for the unadjusted treatment effect parameter to find an approximation for the difference between the unadjusted and adjusted treatment effect parameters. From this it is shown that the treatment effect parameters coincide if the function linking the linear part of the model to the expected value of the outcome is linear or exponential. Finally, the authors show that these results hold for the equivalent GLMs using the maximum likelihood estimation equations. The details of these methods are given in Appendix C.1.

The method of Neuhaus & Jewell [7] is referred to as a geometric approach and uses the relationship between the expected value of the outcome in the control treatment arm and the absolute difference between expected values of the outcome in each treatment arm. Having defined the appropriate GLMs, the authors express the absolute difference between expected value of the outcome in each treatment arm in terms of the model parameters. In particular, they consider for the adjusted model the difference between expected outcome in the treatment arms as a function of the expected outcome in the control treatment arm. For example, the graphs in Figure 4.1 (page 65) plot the difference between arms (Δ) over the expected value in the control treatment arm (μ_0), for three values of the treatment effect parameter (β) for logit, identity, and log link functions. Averaging this curve with respect to the distribu-

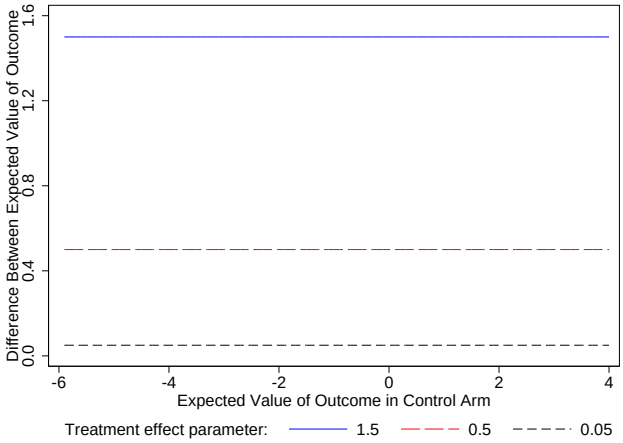
tion of the covariate gives a point on that (μ_0,Δ) plane, which is the expected value for the outcome in the control arm and absolute difference between arms under the unadjusted model. Using Jensen's inequality, the authors show that the unadjusted and adjusted treatment effect parameters coincide when there is no treatment effect or the GLM link function is linear or log, but the adjusted treatment effect is greater in magnitude when using a logistic link. That is, the adjusted log odds ratio will be further from one. Finally, Neuhaus & Jewell [7] use a first-order Taylor expansion to give an approximation for the unadjusted treatment effect parameter in terms of the adjusted treatment effect parameter and covariate distribution. The details of these methods are given in Appendix C.2.

There are some notable differences between these two papers. Gail et al. [6] rely on a second order Taylor series approximation to obtain an expression for the unadjusted treatment effect parameter in term of the adjusted treatment effect parameter. Whilst this generates legitimate conclusions, it gives little insight into the reason for the asymptotic bias or the direction of the bias, or to the validity of conclusions where the Taylor series is not an appropriate approximation. The method of Neuhaus & Jewell [7], however, does not rely on Taylor series to assess the direction of asymptotic bias or to establish conditions in which there is no asymptotic bias. This approach gives conclusions that are valid where the Taylor series approximation may not hold. It also gives a more intuitive and insightful consideration of the mathematical reason for, and direction of, an asymptotic bias between the unadjusted and adjusted treatment effect parameters. Also note that the expression given by Gail et al. [6] for the asymptotic bias and the expression given by Neuhaus & Jewell [7] are not directly comparable as they are Taylor series approximations taken at different points (at covariate effect equal to zero and treatment effect equal to zero, respectively).

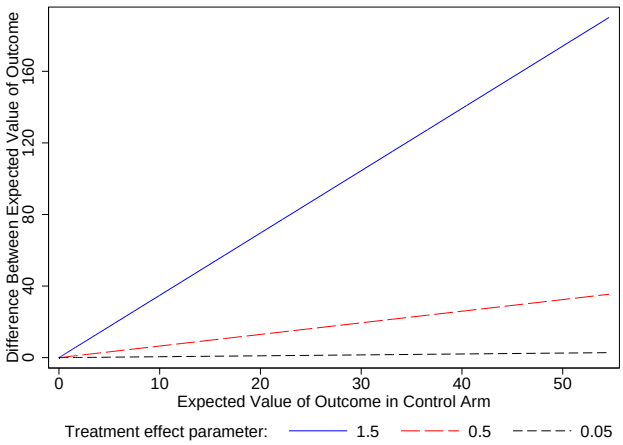
For mixed effects models, Raab & Butcher [23] give an expression for the change in estimate of treatment effect when adjusting for covariates in linear mixed effects models. This follows from the analysis of covariance estimates of the parameters and is given in Section 3.1.1.



(a) With logit link function



(b) With identity link function



(c) With log link function

Figure 4.1: Difference between expected values of the outcome in each treatment arm, Δ, plotted against the expected value of the outcome in the control treatment arm, μ_0. For treatment effect parameters, $\beta = 1.5, 0.5, 0.05$.

4.1.2 Precision of treatment effect estimate

Robinson & Jewell [3] investigate the effect on the precision of estimates of treatment effect when adjusting for a binary covariate in a logistic regression model. This work is then extended to GLMs with a binary outcome by Robinson et al. [5]. They find conditions on the link function for which there is greater precision under the unadjusted or adjusted models. The authors express the treatment effect parameters in terms of conditional probabilities. Using standard likelihood theory results, the variances for treatment effect estimators are then given in terms of these probabilities and sample sizes. Using Jensen's inequality, Robinson et al. [5] show that if the link function is linear or log, the adjusted model gives greater precision of the treatment effect estimate. Also, for other link functions such as logit and probit, the adjusted treatment effect estimator has less precision.

Raudenbush [24], Murray [16] and Raab & Butcher [23] consider the analysis of CRTs with continuous outcome measures, analysed with a linear mixed effects model, and give some analytic results for the precision of treatment effect estimates. They give expressions for the variances of the unadjusted and adjusted treatment effect estimates. Comparing these expressions, they conclude that reducing the residual cluster level and individual level variances of the outcome through covariate adjustment can improve the precision.

4.1.3 Power of test of no treatment effect

As described in Section 2.1.3, the power of tests of no treatment effect can be compared by their asymptotic relative efficiency (ARE). This was introduced in Section 2.1.3. Recall, we have a parameter of interest θ and we wish to test $\theta = \theta_0$. We also have a statistic T that is a consistent estimator of $\mu_T(\theta)$, a monotone function (at least close to $\theta = \theta_0$) of θ. A definition of ARE called the Pitman efficiency of two such statistics T_1 and T_2 that are asymptotically normally distributed with mean $\mu_T(\theta)$ and variance $\sigma^2_T(\theta)$is given by [28]:

$$ARE(T_1 \text{ to } T_2 \text{ at } \theta = \theta_0) = \left[\frac{\mu'_{T_1}(\theta_0)}{\mu'_{T_2}(\theta_0)}\right]^2 \left[\frac{\sigma^2_{T_2}(\theta_0)}{\sigma^2_{T_1}(\theta_0)}\right]$$

We can test the null hypothesis $\beta = 0$ by the adjusted or unadjusted Wald test statistics

$$\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})}} \quad \text{or} \quad \frac{\hat{\beta}^*}{\sqrt{var(\hat{\beta}^*)}}$$

which have asymptotic normal distributions with means $\beta/\sqrt{var(\hat{\beta})}$ and $\beta^*/\sqrt{var(\hat{\beta}^*)}$ and variances 1.

Four papers [3–5, 58] evaluate the ARE of tests of no treatment effect based on the unadjusted and adjusted estimators. There is a progression from specific models to general classes of model. Begg & Lagakos [58] consider only the logistic regression model. Robinson & Jewell [3] consider the logistic regression model with a binary covariate; this method is extended to GLMs with a binary outcome and covariate by Robinson et al. [5]. The method and results of Neuhaus [4] consider all GLMs.

Robinson et al. [5] express the ARE in terms of the GLM link function and conditional probabilities for the outcome variable. By the Cauchy-Schwarz and Jensen's inequalities, they show that the ARE of the adjusted model to the unadjusted model is greater than or equal to one. That is, the adjusted model is more efficient and has greater power for a test of no treatment effect. Details of this method are given in Appendix C.4. Neuhaus [4] use expressions for the estimator variances in terms of expected value then, by application of the Cauchy-Schwarz inequality, conclude the same results as Robinson et al. [5] but for all GLMs. Details of this method are given in Appendix C.5.

The Optimal Design software documentation [26] and Konstantopoulos [25] present analytic results for the power of unadjusted and adjusted analyses of CRTs using linear mixed effects models. The Optimal Design software documentation [26] considers a nested ANOVA, which is equivalent to fitting a linear mixed effects model using restricted maximum likelihood. Power can be calculated by comparing central and non-central F distributions. Konstantopoulos [25] has an equivalent approach but in terms of t distributions. These methods and results are summarised in Section 3.1.2.

4.1.4 Review of analytic methods

There are several strengths to the analytic methods described in this section. The major advantage is the ability to generate absolute conclusions. Results can be found that apply to all link functions, or to all values of parameters (such as intercept, treatment effect, or covariate effect). Such universal conclusions allow us to confidently use these results without fear of special cases or quirks in particular models; special cases do of course exist, but they are uncovered by the analytic methods rather than remaining undiscovered. Analytic methods also allow us to establish necessary and sufficient conditions for particular results. For example, Gail et al. [6] establish that there is no asymptotic bias *if and only if* the link function is linear or log. Therefore, we can be absolutely confident that under the conditions of the given theorem there is no asymptotic bias when using these link functions, but that for any other link function there will be an asymptotic bias. A final advantage of analytic methods is that they can provide insight into the mathematical workings of models and explicitly

show the role of any assumptions. For example, the geometric approach of Neuhaus and Jewell [7] gives some insight into the reasons for asymptotic bias under certain models.

However, analytic methods have some limitations. Firstly, the methods described here only consider the straightforward case of adjusting for a single covariate (or vector of covariates). Extension of these methods to more general cases of adjusting for any subset of a selection of covariates may not be straightforward at all. The often strong assumptions on which these analytic methods are based also restricts this approach. Some results of analytic methods (for example, the expression for the magnitude of asymptotic bias) rely on approximations such as the second-order Taylor series. Where the conditions of the approximation do not hold, such as when the parameter is not close to zero, these results are not reliable. Further, these expressions are only approximations when we truncate the Taylor series, rather than exact results. Therefore, such results are quite limited and cannot be immediately generalised to all situations. Finally, a number of results rely on asymptotic properties of estimators. In small trials, asymptotic results may not be appropriate and so our conclusions may not be relevant.

These disadvantages can limit the utility of analytic methods in investigating the effects of covariate adjustment. If we wish to investigate scenarios where the approximations, asymptotic results or assumptions used in these methods are not appropriate, then our analytical methods may well break down. For example, as noted by Neuhaus & Jewell [7], the integration necessary to evaluate the magnitude of asymptotic bias is often intractable, so an approximation such as the Taylor series is required. Further, analytic methods may not be useful if we wish to investigate more complex models such as introducing further covariates, or compare multiple models or assumptions. Analytic methods may simply fail to give useful results when more relationships are introduced. We may also find that the analysis diverges into multiple separate analyses for each combination of relationships between variables, or set of assumptions. While theoretically analytical results could be found, this multiplicity will not help to give overall conclusions. Finally, if we wish to investigate a particular policy of covariate adjustment rather than inclusion or exclusion of a single covariate, then analytic methods are limited.

4.2 Simulation

Simulation studies can be used to investigate the sampling distribution of an estimator or test statistic. In terms of the effects of covariate adjustment on asymptotic bias,

precision and efficiency, we wish to investigate the sampling distribution of treatment effect estimators and related test statistics. Simulation allows us to approximate sampling distributions under chosen conditions (the underlying data generating model) without requiring analytic investigation.

A simulation is carried out by:

1. Specifying the data generating model.
2. Producing independent data sets from the data generating model.
3. Calculating the estimator or test statistic of interest for each data set.
4. Calculating summary statistics of the sample of the estimator or test statistic, to estimate true sampling distribution parameters or properties of interest.

With a large number of simulated data sets, the summary statistics are good approximations to true sampling distribution parameters. We can improve the precision with which the true parameters are estimated by increasing the number of simulated data sets.

In the models we are considering, the complete data generating model can be considered in three parts:

1. The model of outcome on predictor variables (that is, treatment allocation and covariate).
2. The parameter values for that model.
3. The distribution of covariates and treatment allocation.

One of three methods is used to specify these three factors in each simulation study. In the first, all three parts are specified analytically and then all data is generated directly from probability distributions. For example, Negassa & Hanley [27] specify a logistic model with parameter values. The covariate is binary with equal probability of either value, or continuous and normally distributed. The expected value of the outcome is fixed for the control treatment arm at the lower level of the binary covariate (or mean of the continuous covariate). The second method is to use data from one or more RCTs for baseline covariate, and possibly treatment allocation, information. The model and parameter values are then chosen by the investigator, and outcome data is produced by simulation. For example, Hernandez et al. [59] use baseline data from seven RCTs in traumatic brain injury. Missing covariate data was imputed, then outcome data was simulated with a chosen treatment effect. The final option is

to sample, with replacement, all data from existing trial data, so data are simulated without reference to a standard distribution. For example, Hernandez et al. [29] sample with replacement from hypothetical trial data sets. The same method could be applied to actual trial data.

In any simulation it is important to choose appropriate parameter values and variable distributions, except where trial data is sampled. Multiple values of a parameter are typically investigated in any one simulation study, to consider the effect the parameter value has on outcome, if any at all. Values for parameters and variable distributions are chosen on the basis of previous knowledge, for example from reviews or meta-analyses of trials, from particular trials, or on the basis of wider knowledge. When parameter values are varied, suitable values can be found in the same way. For example, Hernandez et al. [29] consider treatment effect values across the range observed in a published survey of thirteen cardiovascular RCTs. In addition, distributions for covariates are typically chosen to be common and easy to generate from, so that standard distributions in software can be used.

An important issue in the simulation studies presented in this section is quantifying the effect of covariate adjustment. In studies investigating bias some papers give the unadjusted estimate, which can be compared directly to the adjusted parameter value, others give the absolute difference between unadjusted and adjusted estimates, while some give the bias as a percentage. Precision of unadjusted and adjusted estimates is not compared in the studies in this section. The effect on test efficiency is reported by some giving the relative efficiency/power of the unadjusted test to the adjusted test, or by giving only the empirical power of the unadjusted test, or in some cases by reporting the percentage reduction in required sample size when adjusting for the covariate.

4.2.1 Simulations used to investigate the effects of covariate adjustment

A number of papers have used simulation to investigate and quantify the effects of covariate adjustment on treatment effect, precision and power in GLMs. As an example, consider one simulation study of Negassa & Hanley [27]. Data was generated from the logistic regression model

$$logit[pr(Y = 1)] = \alpha + \beta X + \gamma Z$$

where Y is a binary outcome variable, X is a binary treatment variable and Z is a binary covariate. The parameters were chosen to give treatment effect with odds ratio of 2.0, covariate effect with odds ratio of 2.0, and $pr(Y = 1|Z = 0) = 0.4$. It was

also assumed that $pr(Z = 1) = pr(Z = 0) = 0.5$ and sample size was 500. Analyses of 10,000 simulated data sets were carried out under the assumptions of the adjusted model above, and the unadjusted model omitting the covariate Z:

$$logit[pr(Y = 1)] = \alpha^* + \beta^* X \ .$$

The simulation estimate of relative asymptotic bias was then calculated by

$$\frac{|\tilde{\hat{\beta}}^* - \tilde{\hat{\beta}}|}{\tilde{\hat{\beta}}}$$

where $\tilde{\hat{\beta}}^*$ is the median of the 10,000 estimated unadjusted treatment effect parameters $\hat{\beta}^*$. And $\tilde{\hat{\beta}}$ is the median of the 10,000 estimated adjusted treatment effect parameters $\hat{\beta}$. The simulation estimate of power under either the adjusted or unadjusted model was given by the proportion of simulated data sets that show a significant (at 5% level) treatment effect.

Five papers have included simulation studies to investigate the effect of covariate adjustment on asymptotic bias [6, 7, 27] and test efficiency [27, 29, 59] in GLMs. Gail et al. [6] simulate the asymptotic bias for a trial with 100 patients. Outcome data is simulated and analysed for six models. A binary covariate is used with two distributions (one symmetric). Three different treatment effects are considered. Neuhaus & Jewell [7] consider asymptotic bias in GLMs with logistic, complementary log-log, and probit link functions. Data for two independent covariates is generated from a bivariate normal distribution, so this is not strictly a simulation of trial data as one of these continuous covariates is taking the role of treatment arm. Each data set contains 500 observations. The simulation studies are repeated with 100 observations and gamma distributed covariates, but results are not given. Neuhaus & Jewell [7] report the observed unadjusted treatment effect, which can be compared to the unadjusted treatment parameter. Two values for the effect of the omitted covariate are considered. Negassa & Hanley [27] report relative asymptotic bias and power for a logistic regression model. The treatment effect parameter was fixed throughout (OR=2.0) while the covariate effect was varied (OR between 2.0 and 10.0). Both a binary (with symmetric distribution) and continuous (normally distributed with mean zero and variances 1.0 and 5.0) covariate are considered. Sample sizes of 500 and 1000 are used. Results are reported as line graphs of percentage asymptotic bias or power against covariate effect. Hernandez et al. [29] consider efficiency for a logistic regression model with a binary covariate. Data is generated by sampling with replacement from hypothetical trial data sets of 360 patients. Four values for covariate effect are considered (unadjusted OR=1, 2, 5 and 10) and three values of treatment effect (unadjusted OR=1, 1.4, and 1.7). Outcome incidence and covariate prevalence are also varied with sample size increased for small values of both, to prevent zeroes in denominators on odds ratios.

The power of adjusted and unadjusted tests, and percentage reduction in required sample size when adjusting, are reported for each simulation. Hernandez et al. [59] consider test efficiency for a logistic regression model. Baseline data is taken from seven large trials in traumatic brain injury, and outcome data is simulated using a full model containing seven covariates. Unadjusted analysis (no covariates included) is compared to three adjusted analyses: the first including a single continuous covariate, the second including two more covariates, and the last being the full model. Results are reported as the percentage reduction in required sample size compared to the unadjusted analysis.

Simulations have also been used in investigations of the effects of covariate adjustment in the analysis of CRTs. Nixon & Thompson [50] investigate the effect on the precision of treatment effect estimates when adjusting for baseline in the analysis of repeated cross-sectional CRTs with binary outcomes. Austin [51] investigates the power of methods of adjusting for baseline in the analysis of repeated cross-sectional CRTs with binary outcomes. Austin [51] applies a total of nine models to simulated trial data. Two simulation studies [54, 55] have been carried out to investigate the effects of incorporating separate individual and cluster level covariate effects in the analysis of CRTs with linear mixed effects models.

4.2.2 Review of simulation methods

Using simulation studies to investigate the effects of covariate adjustment in RCTs addresses some of the limitations of a purely analytic approach. Conclusions from simulation studies are not asymptotic results, and do not require particular approximations. When performing simulation studies, it is straightforward to vary parameter values, covariate distributions and investigate a variety of effects. It is also easier to investigate more complex covariate situations, different algorithms for choosing covariates, and different analysis methods. Simulation studies also quantify the effects being investigated, which can be reported in a variety of useful ways (for example, percentage sample size reduction). The precision of the results provided by simulation depends on the sample size, or number of repetitions, of the simulation. The precision of results can be increased by increasing the simulation sample size.

Simulation methods do have some weaknesses. Firstly, a number of features, namely the model, parameter values, covariate distributions, and sample sizes, must be specified to generate simulated data. This may be done implicitly by choosing trial data from which to sample. Conclusions are dependent on these choices and so are not easily generalisable. No absolute conclusions, or conditions, are found as in some analytic methods. Practically, simulation studies may also be costly to run in terms of

computation and time. Increasing the number of factors that are varied (for example, parameter values or distributions of variables) increases the number of separate simulations to be run. Improving the precision of results also requires large increases in the number of simulation iterations to be run, increasing the time required.

4.3 Conclusion

Both analytic and simulation methods have been used to investigate the effects of covariate adjustment in the analysis of individually randomised and cluster randomised trials. Different analytic methods are used for considering each effect (asymptotic bias, precision, test efficiency) and models, whereas simulation provides a unified approach to investigating and quantifying effects. The advantages of an analytic approach lie in the ability to prove absolute conclusions and theorems, but are limited by relying on asymptotic results and specific assumptions. Simulation studies can be used where analytic approaches are unsuitable, and are a more flexible and practical method for quantifying the effects of covariate adjustment.

Analytic and simulation methods will both be useful in extending the investigation of the effects of covariate adjustment in the analysis of CRTs. Simulation in particular provides a powerful tool to investigate the behaviour of mixed effects models under a variety of assumptions and conditions.

Chapter 5

Review of the use of covariates in cluster randomised trials

The aim of this thesis is to provide practical guidance for choosing covariates in the analysis of CRTs, so a consideration of current practice in adjusted analyses and choosing covariates in CRTs is important. In this chapter I present some results from a review of the use of covariates in CRTs using a sample of published reports.

The work reported in this chapter is part of a comprehensive review of the use of covariates in the design, analysis, and reporting of CRTs [60]. This was carried out in collaboration with Noah Ivers, Sandra Eldridge, Monica Taljaard, and Stephen Bremner. The Accepted Author Manuscript of the paper can be found in Appendix F.

5.1 Aims

The aims of the review included assessing adherence to guidance for the use of covariates in randomisation, reporting, and analysis of CRTs. In this chapter I focus only on results pertaining to the use of covariates in analysis. In particular, I wish to address the following aims:

1. To assess the prevalence of adjusted analyses reported for CRTs.
2. To investigate the use of cluster level and individual level covariates in adjusted analyses.
3. To assess adherence to existing guidelines for choosing covariates.

4. To identify if justification is provided for the choice of covariates.

Previous reviews of the use of covariates have excluded CRTs [9, 11, 12, 37, 61]. A review of CRTs in primary care [62] described the use of matching or stratification, and reporting of baseline covariates.

5.2 Methods of the review

Two researchers (Neil Wright and Noah Ivers) independently reviewed all papers. Discrepancies were resolved by discussion. A sample of 15 papers were selected and used in a pilot. Following this, questions were updated to clarify meaning and terminology.

Only analyses of the primary outcome were considered. To be considered, an analysis had to include a comparison between treatment arms, by reporting a treatment effect estimate and standard error or confidence interval, or a P-value. In each CRT report, only the most emphasised or the first reported adjusted analysis (excluding adjusting for baseline measure of outcome or covariates used in randomisation) was used for the purposes of this review. Any method for selecting covariates that used data from the trial under analysis (for example, baseline balance or covariate selection algorithms) is a *post hoc* method.

5.3 Sample of published trial reports used in the review

The sample of 300 CRTs was identified using a published electronic search strategy [63] implemented in Medline. The sample was previously used by Ivers et al. [64] to review the impact of the CONSORT extension to CRTs [65] on the reporting and methodology of CRTs. The following search strategy was used in MEDLINE (reproduced from [64]):

1. randomized controlled trial.pt.
2. animals/
3. humans/
4. 2 NOT (2 AND 3)
5. 1 NOT 4
6. cluster$ adj2 randomi$.tw.
7. ((communit$ adj2 intervention$) OR (communit$ adj2 randomi$)).tw.

8. group$ randomi$.tw.
9. 6 OR 7 OR 8
10. intervention?.tw.
11. cluster analysis/
12. health promotion/
13. program evaluation/
14. health education/
15. 10 OR 11 OR 12 OR 13 OR 14
16. 9 OR 15
17. 16 AND 5

Titles and abstracts were screened in a random order until the target sample size of 300 was reached. Only main reports of CRTs were included. Pilot and feasibility studies, protocols, conference proceedings, and secondary analyses were excluded [64].

The sample includes reports of CRTs published between 2000 and 2008 across 150 English language journals. The journals include general medical journals (for example: the New England Journal of Medicine (NEJM), the Journal of the American Medical Association (JAMA), The Lancet, and the British Medical Journal (BMJ)), and various specialty journals (for example: the British Journal of Psychiatry, Diabetes Care, the International Journal of Cancer, and the Journal of Nutrition).

5.4 Use of adjusted analyses

Out of 300 trial reports, 219 (73.0%) included at least one adjusted analysis of the primary outcome. Of 207 trial reports that reported a baseline measure of the primary outcome, 155 (74.9%) included an analysis adjusting for a baseline measure of the outcome. Of 174 trials that used covariates in randomisation, 30 (17.2%) included an analysis adjusting for all covariates used in randomisation.

There were 140 (46.7%) trials that included an analysis adjusting for covariates other than baseline measure of outcome or covariates used in randomisation. In 50 reports published in journals with impact factor (from journal citation reports, ISI Web of Science, 2009) greater than 10, 22 (44%) included an analysis adjusting for other covariates, and 118 (47%) out of 250 published in other journals.

The proportion of trial reports including an analysis adjusting for other covariates by number of clusters in each treatment arm is given in Table 5.1. The three subgroups were formed by the lower quartile, upper quartile, and middle half of trials. Seven trials are not included as the number of clusters was not reported. The proportion of trial reports including an analysis adjusting for other covariates by overall trial sample size is given in Table 5.2 (page 77).

Clusters in each treatment arm	Number of trials / Relevant trials	(%)
Fewer than 5	33/75	(44%)
5-22	71/147	(48%)
Greater than 22	33/71	(46%)

Table 5.1: Proportion of CRT reports that included an analysis adjusting for covariates other than baseline measure of outcome or covariates used in randomisation, by number of clusters.

Overall trial sample size	Number of trials / Relevant trials	(%)
Fewer than 100 individuals	13/47	(28%)
101-200	26/54	(48%)
201-300	19/29	(66%)
301-400	10/22	(45%)
401-500	6/19	(32%)
501-600	7/13	(54%)
601-700	5/10	(50%)
701-800	6/10	(60%)
801-900	3/7	(43%)
901-1000	3/7	(43%)
Greater than 1000 individuals	36/67	(54%)

Table 5.2: Proportion of CRT reports that included an analysis adjusting for covariates other than baseline measure of outcome or covariates used in randomisation, by total number of individuals in the CRT.

5.5 Choice of covariates in adjusted analyses

Out of 219 trial reports that included an adjusted analysis, in 93 (42.4%) reports authors did not describe when covariates had been chosen. In 71 (32.4%) reports covariates were reportedly chosen *a priori*, but in 55 (25.1%) reports some or all covariates were reported to be chosen *post hoc*. In 73 (33.3%) trial reports, authors gave some justification for the choice of covariates.

One hundred and forty trial reports included an analysis adjusting for covariates other than baseline measure of outcome or covariates used in randomisation. Of these, 7 (5.0%) included only cluster level covariates, while 100 (71.4%) adjusted for only individual level covariates. In 29 (20.7%) analyses both cluster and individual level covariates were included. In four (2.9%) of the trials it was unclear which level of covariates had been included.

When included, the number of cluster level covariates ranged from 1 to 8, with a median of 1. Likewise, the number of individual level covariates included in an adjusted analysis ranged from 1 to 28, with a median of 3. The distribution of cluster level and individual level covariates is shown in Figure 5.1 (page 79).

In trials with fewer than five clusters in each treatment arm, the number of individual level covariates adjusted for ranged from 1 to 14. The range was 1 to 11 covariates for trials with 5 to 22 clusters per arm and for trials with more than 22 clusters per arm. The median and range of the number of individual level covariates adjusted for is given for the total number of individuals in the trial in Table 5.3 (page 80), and a scatter plot is given in Figure 5.2 (page 79) for trial sample sizes up to 1000.

Two trials with fewer than five clusters in each treatment arm included adjusted for one and two cluster level covariates. In nineteen trials with 5 to 22 clusters per arm, the maximum number of cluster level covariates adjusted for was 5. In trials with greater than 22 clusters per arm, up to 8 cluster level covariates were adjusted for.

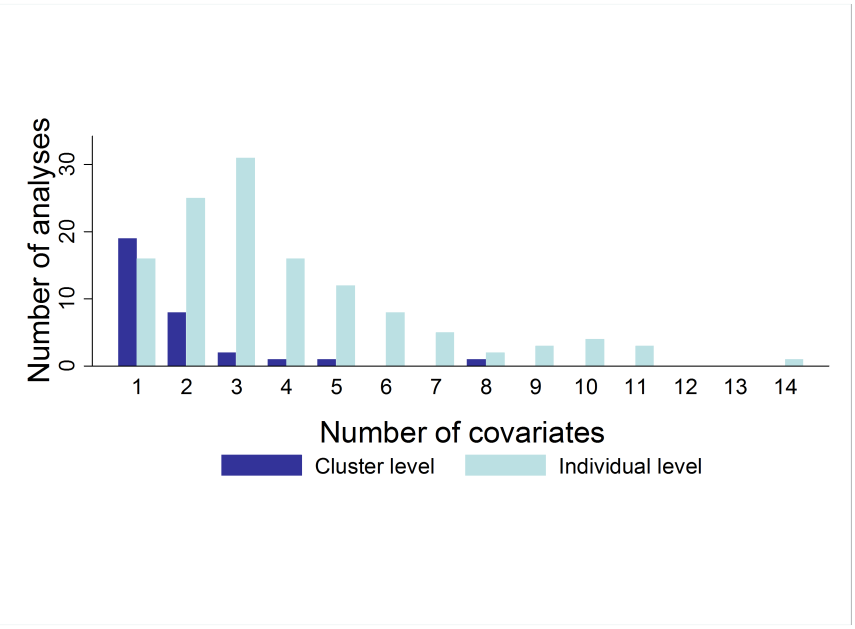


Figure 5.1: Distributions of number of cluster level and individual level covariates included in adjusted analyses.

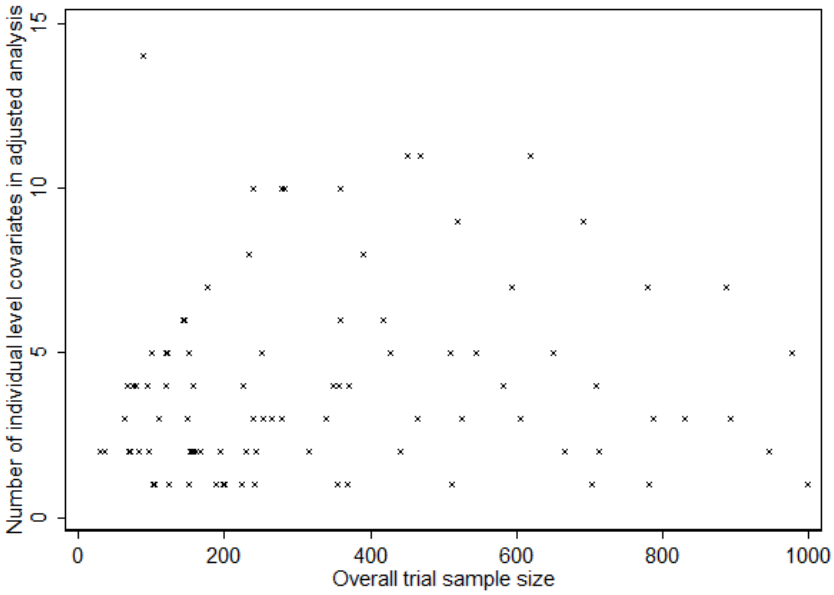


Figure 5.2: Number of individual level covariates included in adjusted analyses against overall trial sample size (excluding trials with sample size greater than 1000).

Overall trial sample size	Number of individual level covariates	
	Median	(min,max)
Fewer than 100	2.5	(2, 14)
101-200	2	(1, 7)
201-300	3	(1, 10)
301-400	4	(1, 10)
401-500	5.5	(2, 11)
501-600	5	(1, 9)
601-700	5	(2, 11)
701-800	2.5	(1, 7)
801-900	3	(3, 7)
901-1000	2	(1, 5)
Greater than 1000	4.5	(2, 28)

Table 5.3: Median (range) of number of individual level covariates (when non-zero) included in adjusted analyses, by total number of individuals in the trial.

5.6 Conclusion

Adjusted analyses are common in reports of CRTs. Further, adjusting for covariates other than baseline measure of outcome and covariates used in randomisation is presented in almost half of the trial reports in the sample. This compares to 34% of trials in the review by Austin et al. [37] which excluded CRTs. This covariate adjustment is observed in CRTs of all sizes and in high impact and other journals. There may be less use of such adjusted analyses in the very smallest trials. The prevalence of adjusted analyses, especially those using covariates other than a baseline measure of outcome, highlights a need for guidance in choosing covariates in CRTs.

Authors of three-quarters of trials reporting a baseline measure of outcome subsequently adjust for it, following recommendations. However, less than one fifth adjust for all covariates used in randomisation. This defies guidance presented in Chapter 2 on adjusting for covariates used in randomisation and can undermine the validity of conclusions.

Adjusting for individual level covariates is more common than adjusting for cluster level covariates. In the smallest trials, fewer individual level covariates are typically adjusted for. Note that we only consider the level at which the covariate is measured, not necessarily the level at which it was included in a model (for example, individual level covariates may have been aggregated at cluster level). The potential utility of adjusting for covariates at each level needs further investigation.

It is often not reported when it was that covariates adjusted for in an analysis were chosen. This prevents readers from assessing the validity of covariate selection and analysis methods. Although this occurs less than in the individually randomised trials included in reviews by Austin et al. [37] and Yu et al. [11]. When this is reported, many covariates are stated to have been chosen *post hoc*. This is inappropriate in a primary analysis of a trial [13, 23]. Guidance for choosing covariates in CRTs should emphasise the importance of choosing covariates *a priori*, just as for individually randomised trials.

In this chapter, I have investigated the use of adjusted analyses and choice of covariates in a wide sample of published CRT reports. With a large proportion of CRTs including adjusted analyses there is a demand for further guidance on choosing covariates in CRTs. Additionally, there is a need for emphasis of following existing guidance for choosing covariates in both individually randomised trials and CRTs.

Chapter 6

Exploration of empirical trial data

In this chapter I present analyses of data sets from two CRTs: the OPERA [66, 67] and FIAT [68, 69] trials. Analyses of data from these example CRTs will allow me to demonstrate known effects of covariate adjustment, and help to identify areas for further investigation. The results will be used to inform the design of simulation studies. The aims of these analyses are:

1. To demonstrate the effects of covariate adjustment on estimated standard error of treatment effect estimates, and estimates of residual variance.
2. To identify hypotheses to be further investigated through analytic work and simulations, with respect to the effects of covariate adjustment in the analysis of CRTs.
3. To provide parameter values and bounds for the simulations of CRT data.

I present estimated ICCs for outcome variables and covariates. I compare results between unadjusted analyses, and analyses adjusting for covariates or covariates aggregated at cluster level. I also compare results between models including a single covariate effect parameter and models with separate within-cluster and contextual covariate effect parameter. In Sections 6.1 and 6.2 I explain the methods used in the analyses and introduce the two data sets. In Section 6.3 I then present results and identify areas for further investigation.

6.1 Methods

6.1.1 Estimating ICCs of outcome variables and covariates

I estimated ICCs for all outcome variables and covariates, giving separate estimates of the ICC for each treatment arm. For this, I used a user-written (by Dr. Obioha Ukoumunne) adaptation of the loneway command in Stata which allows negative ICCs to be estimated. For continuous variables, I also estimated the ICC by fitting linear mixed effects models using restricted maximum likelihood estimation. The variance estimates given by the model were then used to calculate the estimated ICC. Therefore, this estimate of ICC must be non-negative. The linear mixed effects models included a term for treatment arm, so that a single ICC is estimated for both treatment arms. For binary variables, I estimated the ICC from a logistic mixed effects model fitted using maximum likelihood estimation. For binary variables, the ICC on the log odds scale can be defined by [20]

$$\rho_l = \frac{\tau^2}{\tau^2 + \pi^2/3}$$

where τ^2 is the cluster level variance. The individual level variance is given by the variance of the logistic distribution, and does not depend on the cluster prevalence. This definition of the ICC assumes that the binary variable is a dichotomisation of some underlying continuous variable. This assumption may not be suitable for many types of binary variable, such as sex or ethnicity. I estimated the ICC of binary variables using this formula, including the cluster level variance estimate from a logistic mixed effects model. The ICC on the log odds scale is not directly comparable with the ICC on the linear scale. Corresponding ICCs depend on the overall prevalence; some comparison values are given by Eldridge et al. [20]. For binary variables, it is not possible to define an (common across clusters) ICC on the linear scale when using a hierarchical model in the presence of a covariate (such as treatment arm) [20].

6.1.2 Effects of covariate adjustment

I compared standard errors and residual variances between models that exclude and include a single covariate. Linear mixed effects models were used when the outcome variable was continuous, logistic mixed effects models when it was binary, and Poisson mixed effects models for count outcomes. The proportion reduction in estimated standard error of the treatment effect estimate (SE) was calculated as

$$\text{Proportion reduction of SE} = \frac{\text{Unadjusted SE} - \text{Adjusted SE}}{\text{Unadjusted SE}} = 1 - \frac{\text{Adjusted SE}}{\text{Unadjusted SE}} .$$

6.1.3 Effects of separate within-cluster and contextual covariate effect parameters

I compared estimates of treatment effect, standard error, variance, and covariate effects between models that use a single parameter for covariate effect, and models that use two parameters for covariate effect. Using this type of model in the analysis of CRTs was introduced in Chapter 3. Instead of a single covariate effect γ, the linear predictor of the model includes two covariate effects: γ_W is the *within-cluster* coefficient of the covariate effect; and γ_C is the *contextual* covariate effect. For example, the linear mixed effects model incorporating two parameters for the covariate effect is

$$Y_{ij} = \alpha + \beta X_j + \gamma_W (Z_{ij} - \bar{Z}_j) + \gamma_C \bar{Z}_j + u_j + e_{ij}$$

where Y_{ij} is the outcome for patient i in cluster j, X_j indicates treatment arm for cluster j, and u_j and e_{ij} are random effects ($u_j \sim N(0, \tau^2)$, $e_{ij} \sim N(0, \sigma^2)$). $\bar{Z}_j$ is the cluster mean of the covariate for cluster j. Models using a single covariate effect parameter assume that the *within-cluster* and *contextual* covariate effect parameters are equal ($\gamma_W = \gamma_C$).

6.2 Introduction of data sets

6.2.1 The OPERA data set

6.2.1.1 The OPERA trial

The OPERA trial [66, 67] was a CRT investigating the effectiveness of exercise in reducing depression in residents of care homes. The clusters of the individuals in the trial were residential care homes. The original trial analysis had three co-primary outcomes defined for three populations: number of depressive symptoms at 12 months in individuals recruited before randomisation; number of depressive symptoms at six months in those depressed at baseline; and prevalence of depression in all residents present at 12 months. Randomisation was stratified by location, and minimised by type of home and size of home. The allocation ratio was planned to be 1 to 1.5. Seventy-eight clusters were randomised, 35 to the intervention arm and 43 to the control arm, with a total of 2,078 eligible residents. 891 residents had usable data at randomisation and 679 residents were present at 12 months follow-up.

The original analysis of the number of depressive symptoms at 12 months used a linear mixed effects model. Analysis was adjusted for covariates used in randomisation and the proportion of residents with moderate to severe cognitive impairment in the home

at baseline (a cluster level covariate), sex, age at the endpoint, on antidepressants at baseline, Short Physical Performance Battery (SPPB) at baseline, and baseline measure of outcome. The analysis of number of depressive symptoms at six months in those depressed at baseline used a linear mixed effects model, with the same covariates. The analysis of prevalence of depression in all residents present at 12 months used a logistic mixed effects model. The same covariates were included for this analysis, except the baseline measure of outcome was the proportion of residents in the cluster who were depressed at baseline.

6.2.1.2 OPERA Analysis population

The population I used for investigation was all individuals present at 12 months and with data for depression status at 12 months, however in each analysis the population may be reduced due to missing data on outcome or covariates. This data set includes 595 individuals in 78 clusters. Cluster size ranges from one to 17.

6.2.1.3 OPERA Outcome variables

In my analysis, I considered two continuous individual level outcome variables, plus dichotomised versions of these outcome to give two binary outcome variables. The first continuous outcome is an individual's score on the Geriatric Depression Scale (*GDS score*). This outcome is dichotomised to give the binary covariate *depression status*. Generally, a GDS score of 5 or greater is indicative of depression. This binary variable is also used in the original trial analyses.

The second continuous outcome is an individual's score on the Mini Mental State Examination, which I call the *cognitive function score* or *CF score*. This outcome is dichotomised to give the binary variable *normal cognitive function*. A CF score of 25 or greater is considered normal cognitive function. These outcome variables are summarised in Table 6.1 (page 86).

6.2.1.4 OPERA Covariates

Four types of covariates were investigated in this analysis: outcome variables measured at baseline; cluster level covariates used in randomisation; other cluster level covariates; and other individual level covariates. The covariates are summarised in Table 6.2 (page 86). Age at baseline was used as a covariate in the analysis presented in this chapter, although the original analysis of the trial used age at endpoint.

Outcome variable	Type	Notes
GDS score	Continuous	0[best] - 15[worst]
Cognitive function score (CF score)	Continuous	0[worst] - 30[best]
Depression status	Binary	GDS score ≥ 5
Normal cognitive function	Binary	CF score ≥ 25

Table 6.1: OPERA data set outcome variables.

Covariate	Type	Notes
Outcome variables measured at baseline:		
GDS score at baseline	Continuous	0[best] - 15[worst]
Cognitive function score at baseline	Continuous	0[worst] - 30[best]
Depression status at baseline	Binary	GDS score ≥ 5
Normal cognitive function at baseline	Binary	CF score ≥ 25
Cluster level covariates used in randomisation:		
Location of home	Binary	
Size of home (> 32 beds)	Binary	
Other cluster level covariates:		
Proportion depressed in home pre randomisation	Continuous	
Mean CF score in home pre randomisation	Continuous	
Other individual level covariates:		
Age at baseline	Continuous	
Sex	Binary	

Table 6.2: OPERA data set covariates.

6.2.2 The FIAT data set

6.2.2.1 The FIAT trial

The FIAT trial [68,69] was a CRT investigating the effectiveness of financial incentives to improve adherence to anti-psychotic depot medication over a one year intervention period. The clusters in the trial were mental health teams operating in the community, with their patients being the individuals in the trial population. The primary outcome of the trial was adherence level, defined as the percentage of depots administered out of the total number scheduled. Secondary outcomes included clinical improvements and the number of hospital admissions. Randomisation was stratified by the deprivation level of the area in which the health teams operated (categorised as either high or low, and based on data from the mental illness needs index score). Seventy-three clusters were randomised, with a total of 141 patients.

The analysis of the primary outcome used a linear mixed effects model, with a random effect for health teams. Data were used from all patients who had at least four months of complete data during the baseline and follow-up periods. The original primary analysis adjusted for MINI (Mental Illness Needs Index) score, average number of weeks between depots according to treatment cycle, and baseline measure of the outcome variable. Data from 123 patients, in 62 clusters were included in the primary analysis of the trial.

6.2.2.2 FIAT Analysis population

The population used in my analysis was all patients randomised, however in each analysis the population may be reduced due to missing data on outcome or covariates. The full data set includes 141 individuals in 73 clusters. Cluster size ranges from one to seven. Thirty five clusters contain only one individual, and twenty two contain two individuals.

6.2.2.3 FIAT Outcome variables

In my analysis I considered seven outcome variables, of which four are continuous, one is binary, and two are count data. *Adherence* (the percentage of depots administered out of the total number scheduled) is a continuous variable, which is dichotomised to give a binary covariate *achieved at least 95% adherence*. Slippage is the percentage of the prescribed time interval that expired before the depot is taken, and is bounded at 0 and 100 percent. The mean *slippage* during the intervention period is a continuous

outcome variable. Clinical improvement is measured by the Clinical Global Impression Scale, giving continuous outcome *CGI score*. Quality of life is rated by the DIALOG scale, giving the outcome *quality of life*. The *number of psychiatric hospital admissions* and the *number of involuntary hospital admissions* are the two count outcome variables. These outcome variables are summarised in Table 6.3 (page 89).

6.2.2.4 FIAT Covariates

Four types of covariates were included in my analysis: outcome variables measured at baseline; the (binary) cluster level covariate used to stratify randomisation; and other individual level covariates. These include continuous, binary, categorical, and count data types. The covariates are summarised in Table 6.4 (page 89).

Outcome variable	Type	Notes
Adherence	Continuous	Percentage of prescribed depots administered
Achieved at least 95% adherence	Binary	Adherence $\geq$ 95%
Slippage	Continuous	
CGI score	Continuous	
Quality of life	Continuous	
Number of psychiatric hospital admissions	Count	
Number of involuntary hospital admissions	Count	

Table 6.3: FIAT data set outcome variables.

Covariate	Type	Notes
Outcome variables measured at baseline:		
Adherence	Continuous	Percentage of depots administered
Achieved at least 95% adherence	Binary	Adherence $\geq$ 95%
Slippage	Continuous	
Quality of life	Continuous	
Number of psychiatric hospital admissions	Count	
Number of involuntary hospital admissions	Count	
Cluster level covariate used in randomisation:		
MINI score category	Binary	
Other individual level covariates:		
Age	Continuous	
Years of formal education	Continuous	
Number of years since illness diagnosed	Continuous	
Average treatment cycle during baseline	Continuous	
Sex	Binary	
Has children	Binary	

Table 6.4: FIAT data set covariates.

6.3 Results

In this section, I firstly present estimated ICCs of the variables in each data set, and then compare results of unadjusted analyses and analyses adjusting for single covariates. Comparison of results using a single or separate covariate effect parameters are at the end of this section.

6.3.1 Estimated ICCs of outcome variables and covariates

The number of observations and estimated ICCs for outcome variables are given in Table 6.5 (page 91) for the OPERA data set, and Table 6.7 (page 92) for the FIAT data set.

In OPERA, *GDS score* outcome has an estimated ICC very close to zero. The estimate ICC for *cognitive function score* outcome is heterogeneous between treatment arms, with an estimated common ICC close to 0.05. *Normal cognitive function* appears to have quite different ICCs in each treatment arm, however there is a large amount of uncertainty in such estimates of ICCs.

In FIAT, the ANOVA ICC estimates for *adherence* and *achieved at least 95% adherence* show heterogeneity between treatment arms. Where a common ICC is assumed, the estimated ICC is large for *adherence*. There is also a difference between estimated ICCs in each treatment arm for *slippage*, and when a common ICC is assumed it is estimated to be very close to zero.

Estimated ICCs for covariates in the OPERA data set are given in Table 6.6 (page 91), and for the FIAT data set in Table 6.8 (page 93). ICC estimates range from close to zero, to large ICCs (~ 0.5). In the OPERA data set, some covariates show a high degree of heterogeneity of ICC between treatment arms. In the FIAT data set most covariates have similar estimated ICCs between arms

Outcome variable	Number of observations	Estimated ICC (ANOVA)		Estimated ICC (model)
		in experimental arm	in control arm	
GDS score	595	−0.0249	0.0049	<0.0001
Cognitive function (CF) score	549	0.0906	0.0258	0.0544
Depression status (binary)	595	−0.0050	0.0322	0.0152†
Normal cognitive function (binary)	549	0.1067	0.0192	0.1193†

Table 6.5: Estimated ICCs of outcome variables in the OPERA data set. From random effects ANOVA for each treatment arm, and from mixed effects model assuming common ICC across arms. († ICC estimated from logistic mixed effects model is on logistic scale.)

Covariate	Number of observations	Estimated ICC (ANOVA)		Estimated ICC (model)
		in experimental arm	in control arm	
Outcomes measured at baseline:				
GDS Score at baseline	585	0.0865	−0.0567	0.0221
Cognitive function score at baseline	571	0.0696	0.0522	0.0561
Depression status at baseline (binary)	585	0.1105	−0.0570	0.0374†
Normal cognitive function at baseline (binary)	571	0.0148	−0.0013	0.0139†
Other individual level covariates:				
Age at baseline	593	0.1231	0.1715	0.1527
Sex (binary)	595	0.0733	0.0065	0.0378†

Table 6.6: Estimated ICCs of covariates in the OPERA data set. From random effects ANOVA for each treatment arm, and from mixed effects model assuming common ICC across arms. († ICC estimated from logistic mixed effects model is on logistic scale.)

Outcome variable	Number of observations	Estimated ICC (ANOVA)		Estimated ICC (model)
		in experimental arm	in control arm	
Adherence	131	−0.0819	0.5975	0.3463
Achieved at least 95% adherence (binary)	131	−0.0401	0.6183	0.0479†
Slippage (mean during intervention)	131	0.1490	−0.0399	<0.0001
Quality of Life	88	−0.0200	−0.0803	<0.0001
CGI score	107	0.3077	0.3409	0.2771
Number of hospital admissions (psychiatric)	138	0.3026	0.2644	0.0739
Number of hospital admissions (involuntary)	141	0.6202	0.2712	0.5203

Table 6.7: Estimated ICCs of outcome variables in FIAT data set. From random effects ANOVA for each treatment arm, and from mixed effects model assuming common ICC across arms. († ICC estimated from logistic mixed effects model is on logistic scale.)

Covariate	Number of observations	Estimated ICC (ANOVA)		Estimated ICC (model)
		in experimental arm	in control arm	
Outcomes measured at baseline:				
Adherence at baseline	127	0.1299	0.1916	0.1455
Achieved at least 95% adherence at baseline (binary)	127	−0.0112	−0.2012	0.1943†
Slippage (mean during baseline)	127	−0.0226	−0.0184	<0.0001
Quality of Life at baseline	85	0.1045	0.4409	0.2970
Number of hospital admissions (psychiatric) at baseline	138	0.2963	0.8109	0.5271
Number of hospital admissions (involuntary) at baseline	138	0.7366	−0.4266	0.4698
Individual level covariates:				
Age	141	0.1655	0.1537	0.1309
Years of formal education	112	0.1000	0.2054	0.1862
Number of years since illness diagnosed	127	0.3936	−0.0548	0.1805
Average treatment cycle during baseline	127	0.0071	−0.3439	0.0170
Sex (binary)	141	−0.2397	−0.0651	<0.0001†
Has children (binary)	138	0.1482	0.2723	0.2378†

Table 6.8: Estimated ICCs of covariates in FIAT data set. From random effects ANOVA for each treatment arm, and from mixed effects model assuming common ICC across arms. († ICC estimated from logistic mixed effects model is on logistic scale.)

6.3.2 Effects of covariate adjustment on the estimated standard error of treatment effect estimates

6.3.2.1 Continuous outcome variables with very small ICCs

Table 6.9 (page 95) presents estimated standard errors and variances for unadjusted and adjusted analyses of *GDS score* (OPERA), which has an estimated ICC very close to zero. Adjusting for baseline measure of outcome shows the largest reduction (24%) in estimated standard error of the treatment effect estimate. This is accompanied by a reduction in the estimated individual level variance, but an increase in the estimated cluster level variance of the outcome. An increase in cluster level variance is unexpected, however this may be explained by the nature of fitting a mixed effects model. The estimated unadjusted cluster level variance is very close to zero (estimate less than 0.00005) and these variances are constrained to be non-negative. The increase then may be an artefact of fitting a model with a particularly small cluster level variance.

Adjusting for *depression status at baseline* (a dichotomised baseline measure of outcome) reduces the estimated standard error of the treatment effect estimate by 17%. When either of these covariates is cluster aggregated, the reduction in standard error when adjusting is much smaller (and further reduced when using a log odds summary of depression status).

Other covariates do not have a marked effect on the estimated standard error of the treatment effect estimate or on the estimated variances of the outcome, except the covariate *sex* when aggregated at cluster level by log odds. Here, there is an increase in the estimated individual level variance of the outcome and an increase in the standard error of the estimated treatment effect of 8%.

Table 6.10 (page 97) summarises the effects of adjusting for single covariates in an analysis of *quality of life* (FIAT), which has estimated ICC very close to zero. Adjusting for a baseline measure of outcome reduces the standard error of the treatment effect estimate by eighty percent, which is not achieved when adjusting for this covariate summarised at cluster level (two percent reduction). Adjusting for *psychiatric hospital admissions at baseline* reduces the standard error by almost five percent; when the covariate is summarised at cluster level, the reduction in standard error is almost four percent.

Covariate adjusted for	Sample size	SE of effect estimate		Proportion reduction of SE	Cluster level variance of outcome		Individual level variance of outcome	
		Unadjusted	Adjusted		Unadjusted	Adjusted	Unadjusted	Adjusted
GDS score at baseline	585	0.2756	0.2099	0.2386	<0.0001	0.0120	11.0940	6.3220
→ Cluster aggregated	585	0.2756	0.2707	0.0177	<0.0001	<0.0001	11.0940	10.6764
Depression status at baseline	585	0.2756	0.2286	0.1706	<0.0001	<0.0001	11.0940	7.6314
→ Cluster aggregated (proportion)	585	0.2756	0.2716	0.0146	<0.0001	<0.0001	11.0940	10.7676
→ Cluster aggregated (log odds)	585	0.2756	0.2743	0.0048	<0.0001	<0.0001	11.0940	10.7778
Centre	595	0.2733	0.2773	−0.0147	<0.0001	<0.0001	11.0972	11.0635
Size	595	0.2733	0.2732	0.0003	<0.0001	<0.0001	11.0972	11.0891
Proportion depressed in home at baseline	595	0.2733	0.2709	0.0085	<0.0001	<0.0001	11.0972	10.8797

Continued on next page.

Table 6.9: Example of the effects of adjusting for a single covariate in a linear mixed effects model, using OPERA data set. Outcome variable is GDS score. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

Continued from previous page.

Covariate adjusted for	Sample size	SE of effect estimate		Proportion reduction of SE	Cluster level variance of outcome		Individual level variance of outcome	
		Unadjusted	Adjusted		Unadjusted	Adjusted	Unadjusted	Adjusted
Mean CF score in home at baseline	595	0.2733	0.2755	−0.0082	<0.0001	<0.0001	11.0972	11.0826
Age	593	0.2731	0.2731	0.0002	<0.0001	<0.0001	11.0501	11.0437
→ Cluster aggregated (mean)	593	0.2731	0.2726	0.0020	<0.0001	<0.0001	11.0501	10.9954
Sex	595	0.2733	0.2733	<0.0001	<0.0001	<0.0001	11.0972	11.0810
→ Cluster aggregated (proportion)	595	0.2733	0.2745	−0.0043	<0.0001	<0.0001	11.0972	11.0968
→ Cluster aggregated (log odds)	595	0.2733	0.2960	−0.0832	<0.0001	<0.0001	11.0972	11.1703

Table 6.9: Example of the effects of adjusting for a single covariate in a linear mixed effects model, using OPERA data set. Outcome variable is GDS score. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

Covariate adjusted for	Sample size	SE of effect estimate Unadjusted	SE of effect estimate Adjusted	Proportion reduction of SE	Cluster level variance of outcome Unadjusted	Cluster level variance of outcome Adjusted	Individual level variance of outcome Unadjusted	Individual level variance of outcome Adjusted
Quality of Life at baseline	60	0.2413	0.2214	0.0827	<0.0001	<0.0001	0.8116	0.6158
→ Cluster aggregated (mean)	60	0.2413	0.2367	0.0191	<0.0001	<0.0001	0.8116	0.6793
Age	88	0.1906	0.1933	−0.0142	<0.0001	<0.0001	0.7397	0.7386
→ Cluster aggregated (mean)	88	0.1906	0.1941	−0.0187	<0.0001	<0.0001	0.7397	0.7347
Achieved 95% adherence	84	0.1861	0.1825	0.0189	<0.0001	<0.0001	0.6676	0.6426
→ Cluster aggregated (proportion)	84	0.1861	0.1859	0.0008	<0.0001	<0.0001	0.6676	0.6664
→ Cluster aggregated (log odds)	8	0.8197	0.9561	−0.1664	<0.0001	<0.0001	1.0078	0.9141
MINI score category	88	0.1906	0.1926	−0.0108	<0.0001	<0.0001	0.7397	0.7379
Psychiatric hospital admissions at baseline	88	0.1906	0.1814	0.0484	<0.0001	<0.0001	0.7397	0.6685
→ Cluster aggregated (mean)	88	0.1906	0.1834	0.0378	<0.0001	<0.0001	0.7397	0.6833

Table 6.10: Example of the effects of adjusting for a single covariate in a linear mixed effects model, using FIAT data set. Outcome variable is quality of life. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

6.3.2.2 Continuous outcome variables with larger ICCs

The effects of adjusting for single covariates in an analysis of *cognitive function score* (OPERA) (ICC = 0.05) are given in Table 6.11 (page 99). Adjusting for baseline measure of outcome shows the largest reduction in estimated standard error of the treatment effect estimate of 48%. With this there is a large reduction in both the cluster and individual level variance estimates. Adjusting for baseline measure of outcome aggregated at cluster level gives a smaller reduction in standard error (20%), with a larger reduction of cluster level variance (to almost zero) but a much smaller reduction of individual level variance. There is a 15% reduction in standard error when adjusting for *mean CF score (of home) at baseline*, with a large reduction in cluster level variance. Adjusting for sex aggregated at cluster level by log odds increases the estimated standard error by 12%, and increases cluster level variance of the outcome.

Table 6.12 (page 101) summarises the effects of adjusting for single covariates in an analysis of *CGI score* (FIAT), which has an ICC of approximately 0.3. Adjusting for a baseline measure of *quality of life* slightly inflates the standard error of the treatment effect estimate, and only reduces cluster and individual level variances very slightly. An increase in standard error is not expected when there is a reduction in residual variances, however here there are only small changes and the increase in standard error may be explained by the reduction in degrees of freedom caused by adding the covariate to the model. This inflation is larger when the covariate is summarised at cluster level, and there is an increase in the adjusted individual level variance estimate compared to the unadjusted estimate. A similar pattern is seen for the covariate *years of formal education*. Adjusting for *achieved at least 95% adherence* reduces the standard error of treatment effect estimate, more so when summarised by cluster proportion. An increase in the individual level variance of outcome is offset by a reduction of the cluster level variance.

Covariate adjusted for	Sample size	SE of effect estimate		Proportion reduction of SE	Cluster level variance of outcome		Individual level variance of outcome	
		Unadjusted	Adjusted		Unadjusted	Adjusted	Unadjusted	Adjusted
Cognitive function score at baseline	533	0.6962	0.3595	0.4837	2.6655	0.4250	43.6222	13.7126
→ Cluster aggregated	533	0.6962	0.5552	0.2025	2.6655	<0.0001	43.6222	39.5394
Normal cognitive function at baseline	533	0.6962	0.5605	0.1949	2.6655	1.6257	43.6222	29.0360
→ Cluster aggregated (proportion)	533	0.6962	0.5691	0.1825	2.6655	<0.0001	43.6222	43.0253
→ Cluster aggregated (log odds)	533	0.6962	0.6320	0.0923	2.6655	0.2219	43.6222	43.0903
Centre	549	0.6788	0.6704	0.0123	2.3233	1.9783	44.2811	44.3278
Size	549	0.6788	0.6741	0.0069	2.3233	2.2195	44.2811	44.2278
Proportion depressed in home at baseline	549	0.6788	0.6693	0.0139	2.3233	2.0878	44.2811	44.2923
Mean CF score in home at baseline	549	0.6788	0.5732	0.1556	2.3233	0.0770	44.2811	43.7581

Continued on next page.

Table 6.11: Example of the effects of adjusting for a single covariate in a linear mixed effects model, using OPERA data set. Outcome variable is cognitive function score. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

Continued from previous page.

Covariate adjusted for	Sample size	SE of effect estimate		Proportion reduction of SE	Cluster level variance of outcome		Individual level variance of outcome	
		Unadjusted	Adjusted		Unadjusted	Adjusted	Unadjusted	Adjusted
Age	548	0.6804	0.6920	−0.0170	2.3467	2.7243	44.3446	43.5540
→ Cluster aggregated	548	0.6804	0.6796	0.0013	2.3467	2.3181	44.3446	44.3611
Sex	549	0.6788	0.6808	−0.0030	2.3233	2.4953	44.2811	43.2876
→ Cluster aggregated (proportion)	549	0.6788	0.6798	−0.0015	2.3233	2.3180	44.2811	44.2538
→ Cluster aggregated (log odds)	549	0.6788	0.7595	−0.1190	2.3233	2.7519	44.2811	44.5047

Table 6.11: Example of the effects of adjusting for a single covariate in a linear mixed effects model, using OPERA data set. Outcome variable is cognitive function score. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

Covariate adjusted for	Sample size	SE of effect estimate		Proportion reduction of SE	Cluster level variance of outcome		Individual level variance of outcome	
		Unadjusted	Adjusted		Unadjusted	Adjusted	Unadjusted	Adjusted
Quality of Life at baseline	67	0.3795	0.3898	−0.0270	0.4352	0.4079	1.4585	1.4331
→ Cluster aggregated (mean)	67	0.3795	0.4020	−0.0593	0.4352	0.3772	1.4585	1.4943
Years of formal education	87	0.3001	0.3017	−0.0053	0.0723	0.0912	1.7692	1.6886
→ Cluster aggregated (mean)	87	0.3001	0.3075	−0.0249	0.0723	0.1579	1.7692	1.5845
Ahieved 95% adherence at baseline	102	0.3312	0.3241	0.0213	0.5190	0.4473	1.5438	1.5779
→ Cluster aggregated (proportion)	102	0.3312	0.3156	0.0469	0.5190	0.3806	1.5438	1.5833
→ Cluster aggregated (log odds)	8	0.6346	0.6614	−0.0422	0.0000	0.0000	0.6042	0.4375
Sex	107	0.3254	0.3217	0.0116	0.5361	0.5342	1.5793	1.5228
→ Cluster aggregated (proportion)	107	0.3254	0.3209	0.0139	0.5361	0.4973	1.5793	1.5830
→ Cluster aggregated (log odds)	50	0.5452	0.5478	−0.0047	0.5125	0.5272	1.6862	1.6706

Table 6.12: Example of the effects of adjusting for a single covariate in a linear mixed effects model, using FIAT data set. Outcome variable is CGI score. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

6.3.2.3 Binary outcome variable with small ICC

Table 6.13 (page 103) summarises the effects of adjusting for single covariates in an analysis of the binary outcome *depression status* (OPERA), which has an ICC on the logistic scale close to 0.02. The estimated standard error of the treatment effect estimate increases by around 20% when adjusting for either *depression status at baseline* or *GDS score at baseline*. Estimated cluster level variance also increases when adjusting for these covariates. This reflects what would be expected in a logistic regression model, where adjusting for a covariate correlated with outcome will increase the standard error of the estimated treatment effect.

Adjusting for *depression status* or *GDS score measured at baselin*e when aggregated at cluster level reduces the cluster level variance to close to zero, but only marginally reduces the estimated standard error.

Adjusting for other covariates marginally changes the estimated standard error of the treatment effect estimate. As with the continuous outcome variables, adjusting for *sex* aggregated at cluster level by log odds increases the estimated standard error.

Covariate adjusted for	Sample size	SE of effect estimate		Proportion reduction of SE	Cluster level variance of outcome	
		Unadjusted	Adjusted		Unadjusted	Adjusted
Depression status at baseline	585	0.1738	0.2102	−0.2090	0.0377	0.1092
→ Cluster aggregated (proportion)	585	0.1738	0.1687	0.0294	0.0377	<0.0001
→ Cluster aggregated (log odds)	585	0.1738	0.1705	0.0194	0.0377	<0.0001
GDS score at baseline	585	0.1738	0.2095	−0.2053	0.0377	0.0582
→ Cluster aggregated	585	0.1738	0.1695	0.0248	0.0377	<0.0001
Centre	595	0.1750	0.1774	−0.0137	0.0509	0.0504
Size	595	0.1750	0.1751	−0.0005	0.0509	0.0511
Proportion depressed in home at baseline	595	0.1750	0.1682	0.0392	0.0509	<0.0001

Continued on next page.

Table 6.13: Example of the effects of adjusting for a single covariate in a logistic mixed effects model, using OPERA data set. Outcome variable is depression status. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

Continued from previous page.

Covariate adjusted for	Sample size	SE of effect estimate Unadjusted	SE of effect estimate Adjusted	Proportion reduction of SE	Cluster level variance of outcome Unadjusted	Cluster level variance of outcome Adjusted
Mean CF score in home at baseline	595	0.1750	0.1719	0.0179	0.0509	0.0238
Age	593	0.1763	0.1717	0.0261	0.0563	0.0294
→ Cluster aggregated	593	0.1763	0.1689	0.0420	0.0563	0.0116
Sex	595	0.1750	0.1750	0.0002	0.0509	0.0501
→ Cluster aggregated (proportion)	595	0.1750	0.1756	−0.0033	0.0509	0.0500
→ Cluster aggregated (log odds)	595	0.1750	0.1915	−0.0940	0.0509	0.0609

Table 6.13: Example of the effects of adjusting for a single covariate in a logistic mixed effects model, using OPERA data set. Outcome variable is depression status. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

6.3.2.4 Binary outcome variable with larger ICC

The effects of adjusting for a single covariate when analysing the outcome *normal cognitive function* with a logistic mixed effects model are summarised in Table 6.14 (page 106). As with the analysis of *depression status*, adjusting for the baseline measures of outcome increases the estimated standard error of the treatment effect estimate (by 24.6% and 37.7%), while also increasing the estimated cluster level variance. Also as before, adjusting for cluster aggregated baseline measures of outcome decreases the estimated standard error of the treatment effect estimate and estimated cluster level variance of the outcome.

Covariate adjusted for	Sample size	SE of effect estimate		Proportion	Cluster level variance of outcome	
		Unadjusted	Adjusted	reduction of SE	Unadjusted	Adjusted
Normal cognitive function at baseline	533	0.2728	0.3400	−0.2461	0.4275	0.6049
→ Cluster aggregated (proportion)	533	0.2728	0.2259	0.1719	0.4275	0.0297
→ Cluster aggregated (log odds)	533	0.2728	0.2345	0.1406	0.4275	0.0262
Cognitive function score at baseline	533	0.2728	0.3756	−0.3769	0.4275	0.6836
→ Cluster aggregated (mean)	533	0.2728	0.2229	0.1829	0.4275	0.0000
Centre	549	0.2745	0.2750	−0.0019	0.4455	0.4358
Size	549	0.2745	0.2745	−0.0001	0.4455	0.4456
Proportion depressed in home at baseline	549	0.2745	0.2746	−0.0006	0.4455	0.4416

Continued on next page.

Table 6.14: Example of the effects of adjusting for a single covariate in a logistic mixed effects model, using OPERA data set. Outcome variable is normal cognitive function status. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

Continued from previous page.

Covariate adjusted for	Sample size	SE of effect estimate Unadjusted	SE of effect estimate Adjusted	Proportion reduction of SE	Cluster level variance of outcome Unadjusted	Cluster level variance of outcome Adjusted
Mean CF score in home at baseline	549	0.2745	0.2680	0.0236	0.4455	0.3739
Age	548	0.2741	0.2794	−0.0192	0.4427	0.4831
→ Cluster aggregated (mean)	548	0.2741	0.2737	0.0017	0.4427	0.4346
Sex	549	0.2745	0.2773	−0.0104	0.4455	0.4555
→ Cluster aggregated (proportion)	549	0.2745	0.2734	0.0038	0.4455	0.4328
→ Cluster aggregated (log odds)	549	0.2745	0.2994	−0.0910	0.4455	0.4618

Table 6.14: Example of the effects of adjusting for a single covariate in a logistic mixed effects model, using OPERA data set. Outcome variable is normal cognitive function status. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

6.3.2.5 Count outcome variable (Number of involuntary hospital admissions)

The outcome *number of involuntary hospital visits* is a count outcome, but when treated as continuous variable had a large estimated ICC of 0.52. Table 6.15 (page 109) summarises the effects of adjusting for some single covariates in an analysis of *number of involuntary hospital visits* using a Poisson mixed effects model. Adjusting for a baseline measure of outcome reduces the standard error of the treatment effect estimate by 14%, and by 15% when the covariate is summarised by cluster mean. The estimated cluster level variance is reduced in each case. Adjusting for *number of psychiatric hospital admissions at baseline*, or *age*, also shows a notable reduction in the standard error, which is slightly larger when the covariate is summarised by cluster mean. Adjusting for *sex* as an individual level covariate inflates the standard error and the cluster level variance of outcome, but decreases the standard error when summarised by cluster proportion.

Covariate adjusted for	Sample size	SE of effect estimate Unadjusted	Adjusted	Proportion reduction of SE	Cluster level variance of outcome Unadjusted	Adjusted
Number of involuntary hospital admissions at baseline	138	0.6571	0.5627	0.1437	3.3418	1.8327
→ Cluster aggregated (mean)	138	0.6571	0.5560	0.1539	3.3418	1.7671
Number of psychiatric hospital admissions at baseline	138	0.6571	0.6110	0.0703	3.3418	2.6192
→ Cluster aggregated (mean)	138	0.6571	0.6095	0.0725	3.3418	2.6324
Quality of Life at baseline	85	0.7252	0.7523	−0.0374	2.0160	2.0482
→ Cluster aggregated (mean)	85	0.7252	0.7685	−0.0597	2.0160	2.0047
Age	141	0.6575	0.6001	0.0872	3.4130	2.5769
→ Cluster aggregated (mean)	141	0.6575	0.5982	0.0901	3.4130	2.4497

Continued on next page.

Table 6.15: Example of the effects of adjusting for a single covariate in a Poisson mixed effects model, using FIAT data set. Outcome variable is number of involuntary hospital visits. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

Continued from previous page.

Covariate adjusted for	Sample size	SE of effect estimate Unadjusted	SE of effect estimate Adjusted	Proportion reduction of SE	Cluster level variance of outcome Unadjusted	Cluster level variance of outcome Adjusted
Number of years since illness diagnosed	127	0.7510	0.7572	−0.0083	3.9284	4.0071
→ Cluster aggregated (mean)	127	0.7510	0.7499	0.0015	3.9284	3.8855
Sex	141	0.6575	0.6908	−0.0506	3.4130	3.9042
→ Cluster aggregated (proportion)	141	0.6575	0.6430	0.0220	3.4130	3.1202
→ Cluster aggregated (log odds)	68	1.1990	1.1845	0.0121	0.0031	0.0043
MINI score category	141	0.6575	0.6257	0.0484	3.4130	2.8766

Table 6.15: Example of the effects of adjusting for a single covariate in a Poisson mixed effects model, using FIAT data set. Outcome variable is number of involuntary hospital visits. (Variances are as estimated in fitting the mixed effects models. SE = Estimated standard error.)

6.3.3 Effects of using separate within-cluster and contextual covariate effect parameters

Tables 6.16 and 6.17 (page 112) compare estimated standard errors of treatment effect of two linear mixed effects analyses: one using a single parameter for the effect of a covariate; and the other using two parameters for covariate effect and separating the covariate into cluster means and individual residuals. Between these two approaches, there is little difference in estimated standard errors for any of the covariates. In some cases there is a marked difference between the contextual and within-cluster covariate effect (for example, with *GDS score at baseline*, and *cognitive function* score as the outcome), but less than 2% decrease in estimated standard error.

		SE of treatment effect		Proportion	Estimates of covariate effects		
Covariate adjusted for	Sample size	Single CE	Separate CEs	reduction of SE	Single	Contextual	Within-cluster
GDS score at baseline	585	0.2099	0.2079	0.0091	0.6783	0.5345	0.7015
Cognitive function score at baseline	571	0.2784	0.2826	-0.0149	0.0408	0.0629	0.0358
Depression status at baseline	585	0.2286	0.2282	0.0016	3.7241	2.9405	3.8637
Normal cognitive function at baseline							
Age	593	0.2731	0.2725	0.0019	0.0103	0.0598	-0.0066
Sex	595	0.2733	0.2742	-0.0036	-0.2997	-0.1159	-0.3353

Table 6.16: Example of the effects of adjusting for a single covariate using either a single covariate effect parameter, or separate within-cluster and contextual covariate effect parameters. Outcome variable is GDS score.

		SE of treatment effect		Proportion	Estimates of covariate effects		
Covariate adjusted for	Sample size	Single CE	Separate CEs	reduction of SE	Single	Contextual	Within-cluster
GDS score at baseline	543	0.6800	0.6715	0.0125	0.1310	0.5445	0.0797
Cognitive function score at baseline	533	0.3595	0.3641	-0.0128	0.9263	0.9503	0.9214
Depression status at baseline	543	0.6806	0.6792	0.0020	1.2691	2.4918	1.1119
Normal cognitive function at baseline							
Age	548	0.6920	0.6794	0.0182	-0.0955	0.0183	-0.1239
Sex	549	0.6808	0.6794	0.0021	2.1838	1.1283	2.3442

Table 6.17: Example of the effects of adjusting for a single covariate using either a single covariate effect parameter, or separate within-cluster and contextual covariate effect parameters. Outcome variable is cognitive function score.

6.4 Conclusion

The results presented in this chapter demonstrate some effects of covariate adjustment in linear and logistic mixed effects models already described in published research. Some results are less clearly related to known effects and suggest areas for further careful investigation of the effects of covariate adjustment in the analysis of CRTs.

Where a continuous outcome variable has a particularly small ICC, it appears that reducing cluster variance may not be important in reducing the estimated standard error of treatment effect. For example, we observe cases where there is reduced standard error despite an increase in cluster level variance. The reduction in standard error when adjusting for an individual level covariate is smaller when the covariate is aggregated to the cluster level. However, when the outcome variable has a large ICC, there is less difference between adjusting for a cluster aggregated covariate or the individual level covariate itself.

For binary outcomes analysed with a logistic mixed effects model, we observe an increase in the standard error of treatment effect estimate when adjusting for a covariate. This is what occurs for logistic regression models (as described in Chapter 2), so we might also expect an increase in power due to a larger estimated treatment effect. However, when adjusting for cluster aggregated covariates, we observe a reduction in the estimated standard error of treatment effect estimates. Further, where the outcome variable has a larger ICC this effect appears to be greater. The effect of adjusting for individual or cluster level covariates in a logistic mixed effects model on treatment effect estimates, standard error, and power warrant further investigation.

For a variety of outcome variables and covariates, we see little increase in precision by using a separate covariate effects model compared to a single covariate effect model. Only where there is a large difference between the within cluster and contextual covariate effects does there appear to be an advantage, which supports the results of Korendijk et al. [54]. This comparison needs to be extended beyond standard error to also consider power and investigate if there are any conditions under which there is a clear advantage in using the more complex model.

The following chapters of this thesis investigate the effects of covariate adjustment in CRTs, including the areas highlighted by these analyses.

Chapter 7

Simulation study methods

Simulation has widely been used to investigate the effects of covariate adjustment in a variety of situations, as discussed in Chapters 2, 3 and 4. In this chapter I describe the simulation methods I used to investigate the effects of covariate adjustment in the analysis of CRTs using mixed effects models. Monte Carlo simulation allows us to estimate quantities such as power, type I error, bias, coverage, and the distributions of model based standard error and empirical standard error.

In simulation studies, data are firstly generated according to a known model, then the data are analysed with the methods of interest. The model used to generate data, including its parameter values, and the model used to analyse the data must be specified [70]. In the next section, I outline the parameter values chosen to be used to generate simulation data. In Section 7.2 I describe the data generating models and how remaining parameter values were calculated. In Section 7.4 I describe the models used to analyse the simulated data. Details of the software used to carry out simulations are given in Section 7.5. A description of how presented results are calculated is given in Section 7.6.

7.1 Parameter values

In this section I describe the parameter values I chose to initially investigate in simulation studies. In particular these are: cluster size; number of clusters; marginal ICCs; marginal prevalence; difference between treatment arms (treatment effect); and covariate effect. These are sufficient to fully specify the models that are outlined in Section 7.2. The necessary values for the parameters in the linear predictor part of the data generating models can be deduced given values for these parameters. How

they are calculated is also described in Section 7.2.

7.1.1 Cluster size and number of clusters

Initial investigation used fixed cluster sizes of 30 (with 20 clusters per treatment arm) and five (with 60 clusters per treatment arm).

7.1.2 ICCs, variance parameters, and prevalence

For continuous outcome variables and covariates, I assumed a marginal normal distribution. A standard normal distribution was used, as any normally distributed variable can be rescaled to a standard normal as follows:

$$X \sim N(\mu, \sigma^2) \implies \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Therefore, the total variance of continuous variables was fixed at one and specifying a marginal ICC implicitly specified the marginal cluster level and individual level variances. Note that while these marginal parameters are fixed, the residual variances will vary as the distribution or effect of any covariate in the model varies. The calculation of residual variance parameters is described in Section 7.2.

For binary variables, expected value and variance are not independent parameters. Instead, I specified marginal expected values (which is the marginal prevalence) in the control treatment arm. The values initially used were 0.5 (giving the maximum variation of a binary variable) and 0.1.

Marginal ICCs were chosen to be 0.0005, 0.005, 0.05, and 0.1 for each variable. Cluster level covariates, where the ICC is one, were also used. For binary variables, these ICCs were used on the outcome scale, and the equivalent ICC on the linear predictor scale of the data generating model were calculated as necessary.

7.1.3 Treatment effect

Treatment effect was specified by the difference between expected values in each treatment arm. Given cluster size, number of clusters, and marginal distribution of the outcome variable I calculated the minimum detectable treatment effect. This is achieved by solving sample size formulae for the treatment effect. For continuous outcome variables, I used the following sample size formula:

$$N = \frac{2(z_{\alpha/2} + z_{\beta})^2(1 + (m-1)\mathrm{ICC}_Y)}{\Delta^2}$$

Where N is the number of individuals in each treatment arm, m is the cluster size, ICC_Y is the outcome ICC, and Δ is the difference in means between treatment arms. The size of the significance test is α and required power is $1-\beta$. From this the minimum detectable effect is:

$$\Delta = \left(\frac{2(z_{\alpha/2} + z_{\beta})^2 (1 + (m-1)\text{ICC}_Y)}{N} \right)^{\frac{1}{2}}$$

This gives two solutions of the same magnitude, one positive and one negative. I used only the positive treatment effect, as the two are equivalent (by recoding the treatment arm variable). For binary outcomes, I used the following sample size formula:

$$N = \frac{(z_{\alpha/2} + z_{\beta})^2 \times [\Pi_1(1-\Pi_1) + \Pi_0(1-\Pi_0)] \times (1 + (m-1)\text{ICC}_Y)}{\Delta^2}$$

Where Π_1 and Π_0 are the prevalence in each treatment arm and $\Delta = \Pi_1 - \Pi_0$. This gives the minimum detectable effect as:

$$\Delta = \frac{-B \pm \left(B^2 - 4AC\right)^{\frac{1}{2}}}{2A} - \Pi_0$$

where

$$A = N(1 + (m-1)\text{ICC}_Y)^{-1} + (z_{\alpha/2} + z_{\beta})^2$$

$$B = -2N(1 + (m-1)\text{ICC}_Y)^{-1}\Pi_0 - (z_{\alpha/2} + z_{\beta})^2$$

$$C = N(1 + (m-1)\text{ICC}_Y)^{-1}\Pi_0^2 - (z_{\alpha/2} + z_{\beta})^2 \Pi_0(1-\Pi_0) \ .$$

This gives two solutions, one negative and one positive, which have different magnitude if the prevalence is not 0.5. I used the positive treatment effect. In the simulations, I used treatment effects equal to this minimum detectable effect, simulating a sufficiently well powered trial. I also used a treatment effect of zero so that type I error could be estimated.

7.1.4 Covariate effect

Covariate effect was specified by the value of the covariate effect parameter in the linear predictor of each model. The magnitude of this parameter is bounded, since residual variance parameters must be non-negative. Equivalently, the correlation between an outcome variable and a covariate at each level is bounded by -1 and 1. The bounds of the covariate effect parameter under each model are given in Section 7.2. I used covariate effect parameters equal to 0, 0.25, 0.5, 0.75, and 0.95 times the maximum positive covariate effect. This covers a wide range of the possible values for the covariate effect parameter.

Some models used different values for separate covariate effects at the cluster and individual levels. The ratios of the between-cluster to the within-cluster covariate effect parameter used were: 2, 1, 0, -0.5, and -0.66. These values are equivalent to "inequality of covariate effects" [54] of 0.333, 0.5, 1, 2, and 3.

7.2 Data generating models

In this section I outline the models used to generate data. I also give details of how model parameter values were calculated, given the parameters outlined in Section 7.1.

7.2.1 Generating continuous outcome and continuous covariate data

The outcome data generating model when both the outcome and covariate are continuous was a linear mixed effects model. The individual level covariate was also generated with a linear mixed effects model. The model was:

$$Y_{ij} = \alpha + \beta X_j + \gamma Z_{ij} + u_j + e_{ij} \tag{7.1}$$

where

$$u_j \sim N(0, \tau^2)$$
$$e_{ij} \sim N(0, \sigma^2)$$

The covariate data generating model was:

$$Z_{ij} = a_j + b_{ij}$$
$$a_j \sim N(0, \tau_z^2)$$
$$b_{ij} \sim N(0, \sigma_z^2)$$

In this model both the outcome Y and covariate Z have normal distributions, and there is a linear relationship between outcome and covariate. X_j is the treatment arm variable, equal to zero for half of the clusters (the control arm) and one for the other half of the clusters (the experimental treatment arm).

Collecting the cluster level and individual level random effects in equation 7.1, the model for the outcome data can be written as:

$$Y_{ij} = \alpha + \beta X_j + \underbrace{(\gamma a_j + u_j)}_{\text{Cluster level random effect}} + \underbrace{(\gamma b_{ij} + e_{ij})}_{\text{Individual level random effect}} \tag{7.2}$$

Since a_j and u_j are independent and normally distributed, their sum is also normally distributed:

$$(\gamma a_j + u_j) \sim N(0, \gamma^2 \tau_z^2 + \tau^2)$$

Similarly, the sum of b_{ij} and eij is normally distributed:

$$(\gamma b_{ij} + e_{ij}) \sim N(0, \gamma^2 \sigma_z^2 + \sigma^2)$$

So the between-cluster and within-cluster variance components of the outcome Y are $(\gamma^2 \tau_z^2 + \tau^2)$ and $(\gamma^2 \sigma_z^2 + \sigma^2)$, respectively. Therefore, the marginal within treatment arm ICC of the outcome Y in terms of the model parameters is

$$\text{ICC}_Y = \frac{\gamma^2 \tau_z^2 + \tau^2}{\gamma^2(\tau_z^2 + \sigma_z^2) + \tau^2 + \sigma^2} \ . \tag{7.3}$$

The marginal ICC of the covariate Z is

$$\text{ICC}_Z = \frac{\tau_z^2}{\tau_z^2 + \sigma_z^2} \ . \tag{7.4}$$

Any normally distributed variable can be rescaled to have variance of one, so both the outcome and covariate were scaled in this way. This standardised the treatment and covariate effect parameter values, and imposes the following restrictions:

$$\gamma^2(\tau_z^2 + \sigma_z^2) + \tau^2 + \sigma^2 = 1 \tag{7.5}$$

$$\tau_z^2 + \sigma_z^2 = 1 \tag{7.6}$$

Given a value for the marginal ICC of the outcome (ICC_Y), the between-cluster and within-cluster variance components are determined implicitly (by equations 7.3 and 7.5):

$$\text{Marginal between-cluster variance of } Y = \gamma^2\tau_z^2 + \tau^2 = \text{ICC}_Y \tag{7.7}$$

$$\text{Marginal within-cluster variance of } Y = \gamma^2\sigma_z^2 + \sigma^2 = 1 - \text{ICC}_Y \tag{7.8}$$

Similarly, given a value for the marginal ICC of the covariate (ICC_Z) the between-cluster and within-cluster variance components are chosen implicitly (by equations 7.4 and 7.6):

$$\text{Marginal between-cluster variance of } Z = \tau_z^2 = \text{ICC}_Z$$

$$\text{Marginal within-cluster variance of } Z = \sigma_z^2 = 1 - \text{ICC}_Z$$

These complete the relationships necessary to calculate the remaining parameter values, given the marginal ICCs for outcome and covariate, and the covariate effect.

In summary, given values for ICC_Y, ICC_Z, and γ, the remaining parameters values were calculated in the following steps:

1. $\tau_z^2 = \text{ICC}_Z$
2. $\sigma_z^2 = 1 - \text{ICC}_Z$
3. $\tau^2 = \text{ICC}_Y - \gamma^2\tau_z^2$ (Using equation 7.7.)
4. $\sigma^2 = (1 - \text{ICC}_Y) - \gamma^2\sigma_z^2$ (Using equation 7.8.)

Additionally, by setting $\alpha = 0$, the outcome has expected value zero in the control treatment arm ($E[Y_{ij}|X_j = 0] = 0$), and expected value β in the experimental treatment arm ($E[Y_{ij}|X_j = 1] = \beta$).

Note that the adjusted variance components (τ^2 and σ^2) must be non-negative, which restricts the values that the covariate effect parameter γ can take:

$$0 \leq \tau^2 = \text{ICC}_Y - \gamma^2\tau_z^2 \qquad \Rightarrow \qquad \gamma^2 \leq \frac{\text{ICC}_Y}{\tau_z^2} = \frac{\text{ICC}_Y}{\text{ICC}_Z}$$

$$\text{and} \qquad 0 \leq \sigma^2 = (1 - \text{ICC}_Y) - \gamma^2\sigma_z^2 \qquad \Rightarrow \qquad \gamma^2 \leq \frac{1 - \text{ICC}_Y}{\sigma_z^2} = \frac{1 - \text{ICC}_Y}{1 - \text{ICC}_Z}$$

So the value of the covariate effect parameter γ is restricted by the inequality:

$$|\gamma| \leq min\left\{\sqrt{\frac{\text{ICC}_Y}{\text{ICC}_Z}}, \sqrt{\frac{1 - \text{ICC}_Y}{1 - \text{ICC}_Z}}\right\}$$

The value of covariate effect parameters in simulation is chosen as a factor of the maximum possible value for the covariate effect paramter. Call this factor the covariate effect factor (CEF), then if $\text{ICC}_Z \leq \text{ICC}_Y$,

$$\gamma = \text{CEF} \times \sqrt{\frac{1 - \text{ICC}_Y}{1 - \text{ICC}_Z}} \ .$$

The covariate effect parameter γ is related to the individual level correlation between the outcome and covariate, r_i, by

$$\gamma = r_i \frac{sd(\gamma b_{ij} + e_{ij})}{sd(b_{ij})} = r_i\sqrt{\frac{1 - \text{ICC}_Y}{1 - \text{ICC}_Z}}$$

$$\therefore \qquad \text{CEF} = r_i$$

So if $\text{ICC}_Z \leq \text{ICC}_Y$, then the covariate effect factor is equal to the individual level correlation between the outcome and covariate.

If $\text{ICC}_Z \geq \text{ICC}_Y$, then

$$\gamma = \text{CEF} \times \sqrt{\frac{\text{ICC}_Y}{\text{ICC}_Z}}$$

$$\gamma = r_c \frac{sd(\gamma a_j + u_j)}{sd(a_j)} = r_c\sqrt{\frac{\text{ICC}_Y}{\text{ICC}_Z}}$$

$$\therefore \qquad \text{CEF} = r_c \ .$$

where r_c is the cluster level correlation between the outcome and covariate. So if $\text{ICC}_Z \geq \text{ICC}_Y$, then the covariate effect factor is equal to the cluster level correlation between the outcome and covariate.

7.2.2 Generating binary outcome and continuous covariate data

When using logistic mixed effects models, I wished to generate data with given values for the marginal proportion and marginal ICC in the control arm. However, it is not as straightforward to calculate model parameter values as for linear mixed effects models. In Section 7.2.2.1 I describe approaches for generating binary outcome data that were considered but not used. In Section 7.2.2.2 I describe the data generating model that was used in simulations.

7.2.2.1 Methods not used

The logistic mixed effects model is used to analyse binary outcome data from CRTs, and is the analysis method I am investigating. However, the logistic relationship between the linear predictor and outcome variable make it difficult to fix the marginal outcome distribution. Consider the logistic mixed effects model

$$Y_{ij} \sim Bernoulli(p_{ij})$$
$$logit(p_{ij}) = \eta_{ij} = \alpha + \beta X_j + \gamma Z_{ij} + u_j$$

where

$$u_j \sim N(0, \tau^2) \ .$$

The model for the continuous covariate is as in Section 7.2.1. Under this model, the marginal expected value of the binary outcome in the control arm is:

$$E[Y_{ij}|X_j = 0] = E[p_{ij}|X_j = 0] = E[logit^{-1}(\alpha + \gamma Z_{ij} + u_j)]$$

And the ICC within the control treatment arm on the scale of the outcome variable can be defined by:

$$\text{ICC} = \frac{var(E[Y_{ij}|X_j = 0, u_j])}{var(Y_{ij}|X_j = 0)}$$

The variances and expected values in these equations do not generally have a closed algebraic form, so cannot be rearranged to provide expressions for parameter values. There is no straightforward algebraic way to calculate model parameter values to ensure that the marginal distribution of the outcome is fixed.

Linear approximations using Taylor series have been used to simplify the relationship between the outcome variable and the linear predictor in logistic regression models, for example by Gail et al. [6] and Neuhaus & Jewell [7]. The Taylor series expansion of $\eta_{ij}(p_{ij})$ at $p_{ij} = \Pi$ is:

$$\eta_{ij} = logit(\Pi) + \frac{p_{ij} - \Pi}{\Pi(1 - \Pi)} + ...$$

From this we can obtain an approximate expression for p_{ij} in terms of η_{ij} and Π, which is linear:

$$p_{ij} \approx \Pi + \Pi(1 - \Pi)(\eta_{ij} - logit(\Pi))$$

Using this linear approximation the marginal expected value of Y and the marginal ICC of outcome in the control arm are given by:

$$\begin{aligned} E[Y_{ij}|X_j = 0] = E[p_{ij}|X_j = 0] &= E[\Pi + \Pi(1 - \Pi)(\eta_{ij} - logit(\Pi))|X_j = 0] \\ &= \Pi + \Pi(1 - \Pi)(E[\eta_{ij}|X_j = 0] - logit(\Pi)) \\ &= \Pi + \Pi(1 - \Pi)(\alpha - logit(\Pi)) \end{aligned}$$

$$\text{ICC}_Y = \Pi(1 - \Pi)(\gamma^2 \tau_z^2 + \tau^2)$$

This would allow us to calculate parameter values given values for $E[Y_{ij}]$ and ICC_Y. However, the linear approximation is poor as we move away from $p_{ij} = \Pi$. Further, it does not take into account the contribution of the individual level variance of the covariate to the expected value or the ICC of the outcome.

I considered an extension of this approach: using piecewise linear approximations to the logistic function, so that the approximation does not diverge from the function as we move away from $p_{ij} = \Pi$. For example, a piecewise linear approximation could be:

$$p_{ij} = \begin{cases} 0 & \eta_{ij} < -2 \\ \frac{1}{2} + \frac{\eta}{4} & -2 \leq \eta_{ij} \leq 2 \\ 1 & \eta_{ij} > 2 \end{cases}$$

This uses the Taylor series linear approximation of η_{ij} (as above) at $p_{ij} = \frac{1}{2}$ between $\eta = -2$ and $\eta = 2$. For $\eta < -2$ and $\eta > 2$, this approximation takes the values of the limits of p_{ij} as $\eta \to -\infty$ and $\eta \to \infty$, respectively. However, using this approximation we again have expressions for the marginal expected values and ICC of the outcome which cannot be easily used to calculate parameter values. For example, the expected value of the outcome in the control treatment arm is:

$$\begin{aligned} E[Y_{ij}|X_j = 0] &= E[p_{ij}|X_j = 0] \\ &= \left(\frac{1}{2} + \frac{E[\eta_{ij}|X_j = 0]}{4}\right) \times Pr(-2 < \eta < 2) + Pr(\eta > 2) \\ &= \left(\frac{1}{2} + \frac{\alpha}{4}\right)\left(\Phi\left(\frac{2-\alpha}{\sqrt{\tau^2+\sigma^2}}\right) - \Phi\left(\frac{-2-\alpha}{\sqrt{\tau^2+\sigma^2}}\right)\right) \\ &\quad + \left(1 - \Phi\left(\frac{2-\alpha}{\sqrt{\tau^2+\sigma^2}}\right)\right) \end{aligned}$$

Instead of the logistic model, a probit mixed effects model could instead be used to generate binary outcome data in simulation studies of CRTs. The probit mixed effects model to use would be:

$$\begin{aligned} Y_{ij} &\sim Bernoulli(p_{ij}) \\ \Phi^{-1}(p_{ij}) = \eta_{ij} &= \alpha + \gamma Z_{ij} + u_j \\ u_j &\sim N(0, \tau^2) \end{aligned}$$

where the continuous covariate is as before, and Φ^{-1} is the inverse cumulative normal distribution function. The marginal expected value of the outcome in the control arm is given by:

$$\begin{aligned} E[Y_{ij}|X_j = 0] = E[p_{ij}|X_j = 0] &= E[\Phi(\eta_{ij})|X_j = 0] \\ &= E[\Phi(\alpha + \gamma Z_{ij} + u_j)] \\ &= \Phi\left(\frac{\alpha}{\sqrt{1 + \gamma^2(\tau_z^2 + \sigma_z^2) + \tau^2}}\right) \end{aligned}$$

This expression for the marginal expected value of the outcome can be rearranged to give an expression for parameter values. The ICC of the outcome variable also has a straightforward relationship with the ICC on the scale of the linear predictor, which is given in equation 7.12. The probit mixed effects model forms the basis of the latent variable approach used in simulations and described in the next section.

7.2.2.2 Latent variable approach

A binary variable can be generated by dichotomising a latent continuous variable [70]. This is the method I used to generate binary outcome data, and the details of my method are described in this section.

Firstly, a continuous latent outcome variable was generated with a linear mixed effects model:

$$Y_{ij}^* = \alpha + \beta X_j + \gamma Z_{ij} + u_j + e_{ij}$$

where

$$u_j \sim N(0, \tau^2)$$
$$e_{ij} \sim N(0, \sigma^2)$$

The continuous covariate Z was generated using:

$$Z_{ij} = a_j + b_{ij}$$
$$a_j \sim N(0, \tau_z^2)$$
$$b_{ij} \sim N(0, \sigma_z^2)$$

Then binary outcome Y was then generated as

$$Y_{ij} = \begin{cases} 1 & \text{if } Y_{ij}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

A binary outcome generated from dichotomising a continuous latent variable in this way is equivalent to a probit mixed effects model formulation [71].

Given this model, the marginal expected value of Y in the control treatment arm is

$$\begin{aligned} E[Y_{ij}|X_j = 0] = Pr(Y_{ij} = 1|X_j = 0) &= Pr(Y_{ij}^* > 0|X_j = 0) \\ &= Pr(\alpha + \gamma Z_{ij} + u_j + e_{ij} > 0) \\ &= \Phi\left(\frac{\alpha}{\sqrt{\gamma^2(\tau_z^2 + \sigma_z^2) + \tau^2 + \sigma^2}}\right) \end{aligned} \tag{7.9}$$

and in the experimental treatment arm is

$$\begin{aligned} E[Y_{ij}|X_j = 1] = Pr(Y_{ij} = 1|X_j = 1) &= Pr(Y_{ij}^* > 0|X_j = 1) \\ &= Pr(\alpha + \beta + \gamma Z_{ij} + u_j + e_{ij} > 0) \\ &= \Phi\left(\frac{\alpha + \beta}{\sqrt{\gamma^2(\tau_z^2 + \sigma_z^2) + \tau^2 + \sigma^2}}\right) . \end{aligned}$$

Where Φ is the Cumulative Distribution Function of the standard normal distribution. So the difference in expected value of the outcome between treatment arms is:

$$\begin{aligned}\Delta &= E[Y_{ij}|X_j = 1] - E[Y_{ij}|X_j = 0] \\ &= \Phi\left(\frac{\alpha+\beta}{\sqrt{\gamma^2(\tau_z^2+\sigma_z^2)+\tau^2+\sigma^2}}\right) - \Phi\left(\frac{\alpha}{\sqrt{\gamma^2(\tau_z^2+\sigma_z^2)+\tau^2+\sigma^2}}\right)\end{aligned} \tag{7.10}$$

The marginal ICC of the outcome in the control treatment arm on the scale of the latent outcome variable is [71]

$$\text{ICC}_{Y^*} = \frac{\gamma^2\tau_z^2+\tau^2}{\gamma^2(\tau_z^2+\sigma_z^2)+\tau^2+\sigma^2} \tag{7.11}$$

The marginal ICC of the outcome in the control treatment arm (ICC_Y) is related to the ICC of the latent outcome variable (ICC_{Y^*}) by:

$$\text{ICC}_Y = \frac{\Phi_2\left(\Phi^{-1}(\Pi_0), \Phi^{-1}(\Pi_0), \text{ICC}_{Y^*}\right) - \Pi_0^2}{\Pi_0(1-\Pi_0)} \tag{7.12}$$

Where Π_0 is the prevalence of the outcome in the control treatment arm, and Φ_2 is the bivariate normal cumulative distribution function.

By restricting the individual level variance of Y^* to be one

$$\gamma^2\sigma_z^2 + \sigma^2 = 1 \tag{7.13}$$

the marginal distribution of Y is given by a probit mixed effects model.

These complete the relationships necessary to calculate remaining parameter values, given the marginal ICCs for outcome and covariate, and the covariate effect. In summary, given values for ICC_Y, ICC_Z, γ, Π_0, and Δ, I calculated other parameter values as follows:

1. $\tau_z^2 = \text{ICC}_Z$
2. $\sigma_z^2 = 1 - \text{ICC}_Z$
3. $\sigma^2 = 1 - \gamma^2\sigma_z^2$ (Using equation 7.13.)
4. Obtain ICC_{Y^*} from ICC_Y using equation 7.12 and a numerical method (see Appendix D).
5. $\tau^2 = \frac{\text{ICC}_{Y^*}}{1-\text{ICC}_{Y^*}} - \gamma^2\tau_z^2$ (Using equation 7.11.)
6. $\alpha = \Phi^{-1}(\Pi_0)\sqrt{\gamma^2\tau_z^2+\tau^2+1}$ (Using equation 7.9.)
7. $\beta = \Phi^{-1}(\Pi_0+\Delta) \times \sqrt{\gamma^2\tau_z^2+\tau^2+1} - \alpha$ (Using equation 7.10.)

Note that the adjusted variance components (τ^2 and σ^2) must be non-negative, which restricts the values that the covariate effect parameter γ can take:

$$\begin{aligned} & 0 \leq \tau^2 = \frac{(\text{ICC}_{Y^*} - 1)\gamma^2\tau_z^2 + \text{ICC}_{Y^*}}{1 - \text{ICC}_{Y^*}} \\ \Rightarrow \qquad & \gamma^2 \leq \frac{\text{ICC}_{Y^*}}{(1 - \text{ICC}_{Y^*})\tau_z^2} = \frac{\text{ICC}_{Y^*}}{(1 - \text{ICC}_{Y^*})\text{ICC}_Z} \\ \text{and} \qquad & 0 \leq \sigma^2 = 1 - \gamma^2\sigma_z^2 \\ \Rightarrow \qquad & \gamma^2 \leq \frac{1}{\sigma_z^2} = \frac{1}{1 - \text{ICC}_Z} \end{aligned}$$

So the value of the covariate effect parameter γ is restricted by the inequality:

$$|\gamma| \leq min\left\{\sqrt{\frac{\text{ICC}_{Y^*}}{(1 - \text{ICC}_{Y^*})\text{ICC}_Z}}, \sqrt{\frac{1}{1 - \text{ICC}_Z}}\right\}$$

7.2.3 Generating continuous outcome and binary covariate data

The outcome data generating model when the outcome was continuous and the covariate was binary was a linear mixed effects model. The model was:

$$Y_{ij} = \alpha + \beta X_j + \gamma Z_{ij} + u_j + e_{ij}$$

where

$$\begin{aligned} u_j &\sim N(0, \tau^2) \\ e_{ij} &\sim N(0, \sigma^2) \end{aligned}$$

The covariate was generated using a Probit mixed effects model:

$$\begin{aligned} Z_{ij} &\sim Bernoulli(p_{ij}) \\ \Phi^{-1}(p_{ij}) &= \alpha_z + b_j \\ b_j &\sim N(0, \tau_z^2) \end{aligned}$$

The expected value of the covariate Z is

$$E[Z_{ij}] = Pr(Z_{ij} = 1) = \Phi\left(\frac{\alpha_z}{\sqrt{\tau_z^2 + 1}}\right) \tag{7.14}$$

The ICC of the covariate is

$$\text{ICC}_Z = \frac{\Phi_2\left(\Phi^{-1}(\Pi_Z), \Phi^{-1}(\Pi_Z), \text{ICC}_{Z^*}\right) - \Pi_z^2}{\Pi_Z(1 - \Pi_Z)} \tag{7.15}$$

where ICC_{Z^*} is the covariate ICC on the scale of the linear predictor:

$$\text{ICC}_{Z^*} = \frac{\tau_z^2}{\tau_z^2 + 1}$$

The marginal expected values and ICC (within treatment arm) of Y_{ij}

$$E[Y_{ij}|X_j = 0] = \alpha + \gamma\Pi_Z$$
$$E[Y_{ij}|X_j = 1] = \alpha + \gamma\Pi_Z + \beta$$
$$\text{ICC}_Y = \frac{\gamma^2 \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z) + \tau^2}{\gamma^2 \times \Pi_Z(1-\Pi_Z) + \tau^2 + \sigma^2}$$

Total variance of outcome restricted to be one:

$$\gamma^2 \times \Pi_Z(1-\Pi_Z) + \tau^2 + \sigma^2 = 1$$

In summary, given values for ICC_Y, ICC_Z, Π_Z, and γ, the remaining parameters values were calculated in the following steps:

1. Obtain ICC_{Z^*} from ICC_Z using equation 7.15 and a numerical method (see Appendix D).
2. $\tau_z^2 = \frac{\text{ICC}_{Z^*}}{1-\text{ICC}_{Z^*}}$
3. $\alpha_z = \Phi^{-1}(\Pi_z)\sqrt{\tau_z^2 + 1}$
4. $\tau^2 = \text{ICC}_Y - \gamma^2 \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z)$
5. $\sigma^2 = (1-\text{ICC}_Y) - \gamma^2 \times (1-\text{ICC}_Z) \times \Pi_Z(1-\Pi_Z)$

Additionally, by setting $\alpha = -\gamma\Pi_Z$, the outcome has mean zero in the control treatment arm ($E[Y_{ij}|X_j = 0] = 0$), and mean β in the experimental treatment arm ($E[Y_{ij}|X_j = 1] = \beta$).

The adjusted variance components (τ^2 and σ^2) must be non-negative, which restricts the values that the treatment effect parameter γ can take:

$$0 \leq \tau^2 = \text{ICC}_Y - \gamma^2 \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z)$$
$$\Rightarrow \quad \gamma^2 \leq \frac{\text{ICC}_Y}{\text{ICC}_Z \times \Pi_Z(1-\Pi_Z)}$$
$$\text{and} \quad 0 \leq \sigma^2 = (1-\text{ICC}_Y) - (1-\text{ICC}_Z) \times \gamma^2 \times \Pi_Z(1-\Pi_Z)$$
$$\Rightarrow \quad \gamma^2 \leq \frac{1-\text{ICC}_Y}{(1-\text{ICC}_Z) \times \Pi_Z(1-\Pi_Z)}$$

So the value of the covariate effect parameter γ is restricted by the inequality:

$$|\gamma| \leq min\left\{\sqrt{\frac{\text{ICC}_Y}{\text{ICC}_Z \times \Pi_Z(1-\Pi_Z)}}, \sqrt{\frac{1-\text{ICC}_Y}{(1-\text{ICC}_Z) \times \Pi_Z(1-\Pi_Z)}}\right\}$$

7.2.4 Generating binary outcome and binary covariate data

The outcome data when both the outcome and covariate are binary was generated using the latent variable approach previously described. The individual level covariate was also generated with a probit mixed effects model. The outcome model was:

$$Y_{ij}^* = \alpha + \beta X_j + \gamma Z_{ij} + u_j + e_{ij}$$

where

$$u_j \sim N(0, \tau^2)$$
$$e_{ij} \sim N(0, \sigma^2)$$

The covariate was generated using a probit mixed effects model:

$$Z_{ij} \sim Bernoulli(p_{ij})$$
$$\Phi^{-1}(p_{ij}) = \alpha_z + b_j$$
$$b_j \sim N(0, \tau_z^2)$$

The expected value of the covariate Z is

$$E[Z_{ij}] = Pr(Z_{ij} = 1) = \Phi\left(\frac{\alpha_z}{\sqrt{\tau_z^2 + 1}}\right) \tag{7.16}$$

The ICC of the covariate is

$$\text{ICC}_Z = \frac{\Phi_2\left(\Phi^{-1}(\Pi_Z), \Phi^{-1}(\Pi_Z), \text{ICC}_{Z^*}\right) - \Pi_z^2}{\Pi_Z(1 - \Pi_Z)} \tag{7.17}$$

where ICC_{Z^*} is the covariate ICC on the scale of the linear predictor:

$$\text{ICC}_{Z^*} = \frac{\tau_z^2}{\tau_z^2 + 1}$$

Then binary outcome Y was generated as

$$Y_{ij} = \begin{cases} 1 & \text{if } Y_{ij}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

Given this model, the marginal expected value of Yin the control treatment arm is

$$\begin{aligned} E[Y_{ij}|X_j = 0] &= Pr(Y_{ij} = 1|X_j = 0) \\ &= Pr(Y_{ij}^* > 0|X_j = 0) \\ &= Pr(\alpha + \gamma Z_{ij} + u_j + e_{ij} > 0) \\ &= \Pi_Z \times \Phi\left(\frac{\alpha + \gamma}{\sqrt{\tau^2 + \sigma^2}}\right) + (1 - \Pi_Z) \times \Phi\left(\frac{\alpha}{\sqrt{\tau^2 + \sigma^2}}\right) \end{aligned}$$

and in the experimental treatment arm is

$$\begin{aligned} E[Y_{ij}|X_j = 1] &= Pr(Y_{ij} = 1|X_j = 1) \\ &= Pr(Y_{ij}^* > 0|X_j = 1) \\ &= Pr(\alpha + \beta + \gamma Z_{ij} + u_j + e_{ij} > 0) \\ &= \Pi_Z \times \Phi\left(\frac{\alpha + \beta + \gamma}{\sqrt{\tau^2 + \sigma^2}}\right) + (1 - \Pi_Z) \times \Phi\left(\frac{\alpha + \beta}{\sqrt{\tau^2 + \sigma^2}}\right) . \end{aligned}$$

So the difference in expected value of the outcome between treatment arms is:

$$\begin{aligned}\Delta &= E[Y_{ij}|X_j = 1] - E[Y_{ij}|X_j = 0] \\ &= \Pi_Z \times \Phi\left(\frac{\alpha+\beta+\gamma}{\sqrt{\tau^2+\sigma^2}}\right) + (1-\Pi_Z) \times \Phi\left(\frac{\alpha+\beta}{\sqrt{\tau^2+\sigma^2}}\right) \\ &\quad - \Pi_Z \times \Phi\left(\frac{\alpha+\gamma}{\sqrt{\tau^2+\sigma^2}}\right) - (1-\Pi_Z) \times \Phi\left(\frac{\alpha}{\tau^2+\sigma^2}\right)\end{aligned} \tag{7.18}$$

The marginal ICC of the outcome (in the control treatment arm) on the scale of the latent outcome variable is

$$\text{ICC}_{Y^*} = \frac{\gamma^2 \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z) + \tau^2}{\gamma^2 \times \Pi_Z(1-\Pi_Z) + \tau^2 + \sigma^2}$$

The marginal ICC of the outcome (in the control treatment arm) is related to the ICC of the latent outcome variable by:

$$\text{ICC}_Y = \frac{\Phi_2\left(\Phi^{-1}(\Pi_0), \Phi^{-1}(\Pi_0), \text{ICC}_{Y^*}\right) - \Pi_0^2}{\Pi_0(1-\Pi_0)} \tag{7.19}$$

By restricting the individual level variance of Y^* to be one

$$\sigma^2 + \gamma^2 \times (1 - \text{ICC}_Z) \times \Pi_Z(1-\Pi_Z) = 1$$

the marginal distribution of Y is given by a probit mixed effects model.

In summary, given values for ICC_0, Π_Y, ICC_Z, Π_Z, γ, and δ, I calculated other parameter values as follows:

1. Obtain ICC_{Z^*} from ICC_Z using equation 7.17 and a numerical method (see Appendix D).
2. $\tau_z^2 = \frac{\text{ICC}_{Z^*}}{1-\text{ICC}_{Z^*}}$
3. $\alpha_z = \Phi^{-1}(\Pi_z)\sqrt{\tau_z^2+1}$
4. Obtain ICC_{Y^*} from ICC_Y using equation 7.19 and a numerical method (see Appendix D).
5. $\sigma^2 = 1 - \gamma^2 \times (1 - \text{ICC}_Z) \times \Pi_Z(1-\Pi_Z)$
6. $\tau^2 = \frac{\text{ICC}_{Y^*}}{1-\text{ICC}_{Y^*}} - \gamma^2 \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z)$
7. Solve for α using numerical method (see Appendix D):

$$\Pi_0 = \Pi_Z \times \Phi\left(\frac{\alpha+\gamma}{\sqrt{\tau^2+\sigma^2}}\right) + (1-\Pi_Z) \times \Phi\left(\frac{\alpha}{\sqrt{\tau^2+\sigma^2}}\right)$$

8. Solve equation 7.18 for β using numerical method (see Appendix D).

The adjusted variance components (τ^2 and σ^2) must be non-negative, which restricts the values that the covariate effect parameter γ can take:

$$\begin{aligned} & 0 \leq \tau^2 = \frac{\text{ICC}_{Y^*}}{1-\text{ICC}_{Y^*}} - \gamma^2 \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z) \\ \Rightarrow \quad & \gamma^2 \leq \frac{\text{ICC}_{Y^*}}{(1-\text{ICC}_{Y^*}) \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z)} \\ \text{and} \quad & 0 \leq \sigma^2 = 1 - \gamma^2 \times (1-\text{ICC}_Z) \times \Pi_Z(1-\Pi_Z) \\ \Rightarrow \quad & \gamma^2 \leq \frac{1}{(1-\text{ICC}_Z) \times \Pi_Z(1-\Pi_Z)} \end{aligned}$$

So the value of the covariate effect parameter γ is restricted by the inequality:

$$|\gamma| \leq min\left\{\sqrt{\frac{\text{ICC}_{Y^*}}{(1-\text{ICC}_{Y^*}) \times \text{ICC}_Z \times \Pi_Z(1-\Pi_Z)}}, \sqrt{\frac{1}{(1-\text{ICC}_Z) \times \Pi_Z(1-\Pi_Z)}}\right\}$$

7.2.5 Generating cluster level covariate data

For cluster level continuous covariates, the methods described above were used with ICC equal to one. For cluster level binary covariates, the following model was used to generate the covariate Z:

$$Z_j \sim Bernoulli(\Pi_z)$$

Then for continuous outcomes, given values for ICC_Y, Π_Z, and γ, the residual variance parameter values were calculated as:

1. $\tau^2 = \text{ICC}_Y - \gamma^2 \times \Pi_Z(1-\Pi_Z)$
2. $\sigma^2 = 1 - \text{ICC}_Y$

For binary outcomes, given values for ICC_Y, Π_Y, Π_Z, γ, and δ, I calculated other parameter values as follows:

1. Obtain ICC_{Y^*} from ICC_Y using equation 7.19 and a numerical method (see Appendix D).
2. $\sigma^2 = 1$
3. $\tau^2 = \frac{\text{ICC}_{Y^*}}{1-\text{ICC}_{Y^*}} - \gamma^2 \times \Pi_Z(1-\Pi_Z)$
4. Solve for α using numerical method (see Appendix D):

$$\Pi_Y = \Pi_Z \times \Phi\left(\frac{\alpha+\gamma}{\sqrt{\tau^2+\sigma^2}}\right) + (1-\Pi_Z) \times \Phi\left(\frac{\alpha}{\sqrt{\tau^2+\sigma^2}}\right)$$

5. Solve equation 7.18 for β using numerical method (see Appendix D).

7.2.6 Separate individual and cluster level covariate effects

In the previous section, a single parameter is used to include the covariate in the data generating model for the outcome. We also wish to investigate the performance of analysis models when the covariate has different effects on the outcome at cluster and individual levels.

7.2.6.1 Continuous outcome and continuous covariate

The outcome data generating model when both the outcome and covariate are continuous was a linear mixed effects model. The individual level covariate was also generated with a linear mixed effects model. The model was:

$$Y_{ij} = \alpha + \beta X_j + \gamma_b a_j + \gamma_w Z_{ij} + u_j + e_{ij}$$

where

$$u_j \sim N(0, \tau^2)$$
$$e_{ij} \sim N(0, \sigma^2) \ .$$

The covariate data generating model was:

$$Z_{ij} = a_j + b_{ij}$$
$$a_j \sim N(0, \tau_z^2)$$
$$b_{ij} \sim N(0, \sigma_z^2)$$

Given this model, the between cluster variance of the outcome Y is

$$(\gamma_b^2 + \gamma_w^2)\tau_z^2 + \tau^2$$

and the within cluster variance of the outcome Y is

$$\gamma_w^2 \sigma_z^2 + \sigma^2 \ .$$

Therefore, the marginal ICC of the outcome Y is

$$\text{ICC}_Y = \frac{(\gamma_b^2 + \gamma_w^2)\tau_z^2 + \tau^2}{(\gamma_b^2 + \gamma_w^2)\tau_z^2 + \tau^2 + \gamma_w^2\sigma_z^2 + \sigma^2} \ . \tag{7.20}$$

So, given values for ICC_Y, ICC_Z, γ_w, and γ_b the remaining parameters values were calculated in the following steps:

1. $\tau_z^2 = \text{ICC}_Z$
2. $\sigma_z^2 = 1 - \text{ICC}_Z$

3. $\tau^2 = \text{ICC}_Y - (\gamma_b^2 + \gamma_w^2)\tau_z^2$

4. $\sigma^2 = (1 - \text{ICC}_Y) - \gamma_w^2\sigma_z^2$

The adjusted variance components (τ^2 and σ^2) must be non-negative, which restricts the values that the covariate effect parameters γ_w and γ_b can take:

$$0 \leq \tau^2 = \text{ICC}_Y - (\gamma_b^2 + \gamma_w^2)\tau_z^2 \quad \Rightarrow \quad (\gamma_b^2 + \gamma_w^2) \leq \frac{\text{ICC}_Y}{\tau_z^2} = \frac{\text{ICC}_Y}{\text{ICC}_Z}$$

$$\text{and} \quad 0 \leq \sigma^2 = (1 - \text{ICC}_Y) - \gamma_w^2\sigma_z^2 \quad \Rightarrow \quad \gamma_w^2 \leq \frac{1 - \text{ICC}_Y}{\sigma_z^2} = \frac{1 - \text{ICC}_Y}{1 - \text{ICC}_Z}$$

7.3 Covariate used to stratify randomisation

To simulate the use of a continuous covariate to stratify randomisation, the cluster mean of the continuous covariate was used. This was dichotomised, with a cut-off of zero, to create two strata. Within each strata, half of the clusters were allocated to the experimental treatment arm and half were allocated to the control treatment arm.

7.4 Data analysis models

Where the outcome variable was continuous, a linear mixed effects model was used to analyse the simulated CRT data. Where the outcome variable was binary, a logistic mixed effects model was used. Linear mixed effects models were fitted using the Stata 13 command `mixed`, and using Restricted Maximum Likelihood (REML). Logistic mixed effects models were fitted using the Stata 13 command `melogit` which uses maximum likelihood (ML).

In each simulation, two or three analysis were to be compared. For example, the comparison is often between an unadjusted model and a model adjusted for a covariate. For each simulated data set, both analysis method were applied as per the "moderately independent" strategy described by Burton et al. [70].

7.4.1 Testing balance of covariates

Simulations were used to investigate the validity of adjusting for a covariate only if it is imbalanced between treatment arms. A two sample t test of cluster means was used to test imbalance. If the P-value of this test was less than the specified α level, then an adjusted mixed effects model was used. Otherwise, the covariate was not used in

the analysis model. The values of α used were 0, 0.005, 0.1, and 1. With α equal to zero, an adjusted analysis is never used, so this is equivalent to an *a priori* choice of the unadjusted model. With α equal to one, an adjusted analysis is always used, so this is equivalent to an *a priori* choice to adjusted for the covariate.

7.4.2 Testing fixed treatment effects in mixed effects models

There are known issues with testing fixed effects in mixed effects models using a Wald test when there are a small number of clusters, due to the use of asymptotic approximations [72]. Alternative approaches include comparing the Wald statistic to a t-distribution and using a degrees of freedom correction, and using empirical sandwich estimates of standard error [72–75].

I calculated P-values for testing treatment effect parameters using REML estimates and the usual Wald test (comparing the Wald statistic to a normal distribution). Using REML estimates, rather than ML, gives less inflated type I errors [73]. I also avoid simulation of CRTs with a very small number fo clusters per treatment arm, to minimise this error.

7.5 Simulation in Stata

I carried out simulation studies using Stata version 13 (see Appendix E for reference). Stata is a powerful and flexible statistical software package, which allows programming of simulations and analysis and presentation of results. Although Stata has a `simulation` command, I did not use this as it did not provide enough flexibility. Simulations were instead written directly using loops and the Stata `postfile` commands, as described by Feiveson [76]. The seed for the Stata pseudorandom number generator was set at the start of each set of simulations. Seeds were chosen as recommended by Stata's guidance [77] and recorded so that simulations could be reproduced.

The number of iterations used in simulations was 5000 for linear mixed effects models and 500 for logistic mixed effects models. Therefore, with a linear mixed effects model an estimated type I error of 5% would have a Monte Carlo standard error of approximately 0.3%, and an estimated power of 80% would have a Monte Carlo standard error of approximately 0.6%. With a logistic mixed effects model, an estimated type I error of 5% would have a Monte Carlo standard error of approximately 1%, and an estimated power of 80% would have a Monte Carlo standard error of approximately 1.8%.

The Stata command `simsum` [78] was used to produce results where possible. Given the simulation sample of treatment effect estimates and standard errors, this command was used to calculate empirical standard error, and estimates of power and relative error of model-based standard errors, and Monte Carlo standard errors of these estimates.

I used the multiple processor version of Stata on high performance computing cluster. This allowed the simulations to be completed faster than if the processing was done on a desktop computer with fewer processors.

7.6 Presentation of results

Estimates of power, standard error, and relative error of model-based standard errors were produced using the Stata command `simsum` [78]. Using this command, power is calculated as the proportion of Monte Carlo samples where there is a significant treatment effect. The Monte Carlo standard error of power is calculated as

$$\sqrt{\frac{P(1-P)}{n}}$$

where P is the estimate of power and n is the number of Monto Carlo samples.

The standard errors given in the results are the empirical standard error of the treatment effect estimate, which is the standard deviation of the treatment effect estimates in the Monte Carlo sample [78]. Using the `simsum` command, the Monte Carlo standard error of empirical standard error is calculated as

$$\sqrt{V/2(n-1)}$$

where V is the sample variance of the estimates of the treatment effect parameter, and n is the number of Monte Carlo samples.

Using `simsum`, the relative error of model-based standard errors are calculated as

$$\frac{\bar{s}}{\text{Empirical Standard Error}} - 1$$

where $\bar{s}$ is the quadratic mean (or root mean square) of the standard errors produced by the Stata mixed effects model commands. The Monte Carlo standard error of relative error is calculated as

$$(\bar{s}/\sqrt{V})\sqrt{V_{s^2}/(4n\bar{s}^4) + 1/2(n-1)}$$

where V_{s^2} is the sample variance of the square of the standard errors produced by the Stata.

The estimated relative asymptotic bias of treatment effect estimates were calculated as

$$\frac{\bar{\beta} - \bar{\beta}^*}{\bar{\beta}^*}$$

where $\bar{\beta}$ is the mean of estimates of the adjusted treatment effect parameter, and $\bar{\beta}^*$ is the mean of estimates of the unadjusted treatment effect parameter.

For linear mixed effects models, power calculated using a non-central F-distribution is also presented. This is calculated as

$$F_{tail}(1, J-3, \lambda, invF(1, J-3, 0.95)) \tag{7.21}$$

where

$$\lambda = \frac{1}{4}\beta^2 J \left(\text{ICC}_Y - \gamma^2 \text{ICC}_Z + \frac{1 - \text{ICC}_Y - \gamma^2(1 - \text{ICC}_Z)}{m} \right)^{-1} .$$

J is the number of clusters in the CRT. m is the number of individuals in each cluster. β is the treatment effect parameter. $invF(d_1, d_2, x)$ is the inverse cumulative F distribution function, with degrees of freedom d_1 and d_2. $F_{tail}(d_1, d_2, \lambda, x)$ is the reverse cumulative (or upper tail) non-central F distribution function, with degrees of freedom d_1 and d_2, and non-centrality parameter λ.

7.7 Conclusion

In this chapter I have described the methods used to conduct the simulation studies to investigate the effects of covariate adjustment in CRTs. In the following chapter I present results from the simulation studies.

Chapter 8

Simulation study results

In this chapter I present results from simulation studies investigating the effects of covariate adjustment in the analysis of CRTs using mixed effects models. The methods used for simulations have been outlined in Chapter 7.

Estimated type I errors and relative errors for analyses adjusting for a covariate only when it is imbalanced at baseline are presented in Section 8.1. This will allow us to assess the application of guidance to not adjust for a covariate only because it is observed to be imbalanced at baseline [13, 14, 38] to CRTs.

Estimated type I errors and relative errors of model-based standard errors for analyses adjusting for covariates used to stratify randomisation are given in Section 8.2. This will allow us to assess the relevance of guidance to adjust for covariates used to stratify randomisation [13–15] to CRTs.

Results of simulations investigating the effects of covariate adjustment when using linear mixed effects models are given in Section 8.3. For the use of logistic mixed effects models results are given in Section 8.4.

Results for models using separate within-cluster and contextual covariate effect parameters are given in Section 8.5. These extend previous simulation studies on the same issue [54, 55].

Some results of adjusting for a covariate aggregated at cluster level are presented in Section 8.6.

8.1 Adjusting for a covariate imbalanced at baseline

Estimated type I errors and estimated relative error of model-based standard errors (for a treatment effect of zero) for analyses adjusting for a covariate only when imbalanced between treatment arms are presented in Tables 8.1 and 8.2 (page 137). These are for CRTs with 50 clusters in each treatment arm and 30 individuals in each cluster. The test for imbalance was a t-test of cluster means of the continuous covariate.

When unadjusted analyses were always used (by using a test of balance at significance level of zero) or adjusted analyses always used (by using a test of balance at significance level of one), type I error was close to the nominal 5%. When a significance level of 0.05 or 0.1 was used to test for baseline imbalance, standard errors of treatment effect estimates were overestimated and type I error was reduced.

Results for an outcome with an ICC of 0.005 and an individual level covariate with an ICC of 0.005 are presented in Table 8.1 (page 137). For significance levels of 0.05 and 0.1 the smallest estimated type I errors were 1.4% and 0.8%, respectively. In these cases the relative error of the model-based standard errors were 14.8% and 24.8%. Estimated type I errors are shown in Figure 8.1 (page 136).

Results for an outcome with an ICC of 0.1 and a cluster level covariate are presented in Table 8.2 (page 137). For significance levels of 0.05 and 0.1 the smallest estimated type I errors were 2.8% and 1.9%, respectively. In these cases the relative error of the model-based standard errors were 9.2% and 15.5%. Estimated type I errors are shown in Figure 8.2 (page 136).

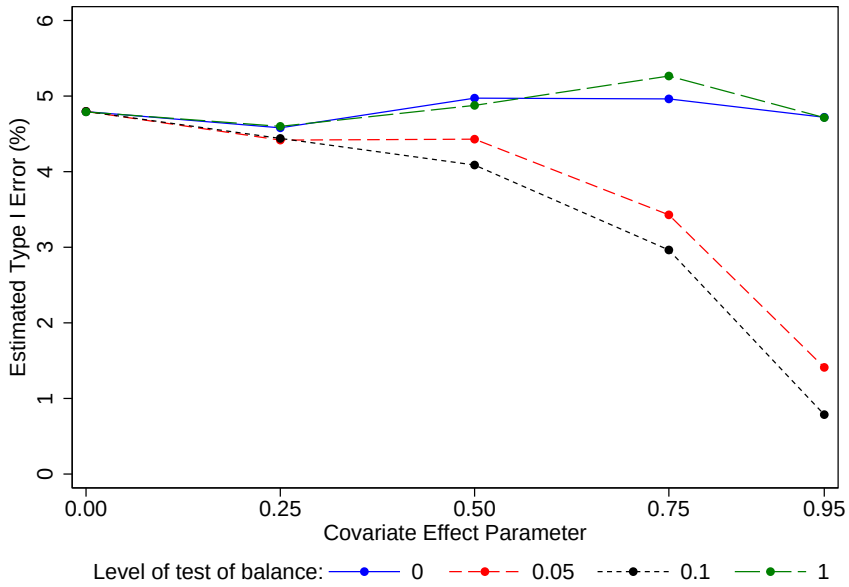


Figure 8.1: Estimated type I error of linear mixed effects model analysis, adjusting for a covariate when a test of balance of covariate is significant. 50 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.005. Covariate ICC = 0.005. Data from Table 8.1.

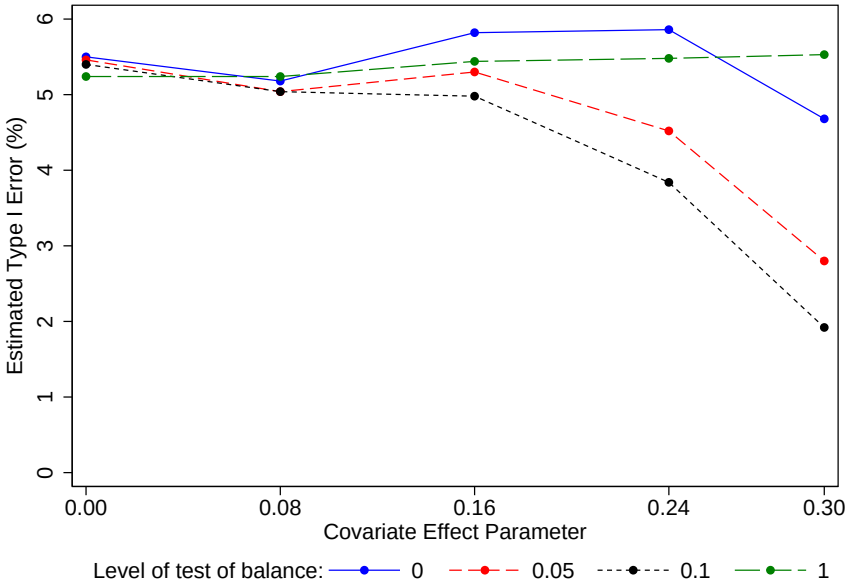


Figure 8.2: Estimated type I error of linear mixed effects model analysis, adjusting for a covariate when a test of balance of covariate is significant. 50 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.1. Cluster level covariate. Data from Table 8.1.

	Estimated Relative Error of SE % (MCSE)				Estimated Type I Error % (MCSE)			
Covariate Effect Parameter	$\alpha = 0$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 0$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 1$
0.00	2.3 (1.0)	2.3 (1.0)	2.3 (1.0)	2.3 (1.0)	4.8 (0.3)	4.8 (0.3)	4.8 (0.3)	4.8 (0.3)
0.25	1.3 (1.0)	2.3 (1.0)	2.6 (1.0)	1.6 (1.0)	4.6 (0.3)	4.4 (0.3)	4.4 (0.3)	4.6 (0.3)
0.50	1.5 (1.0)	4.5 (1.1)	5.9 (1.1)	2.0 (1.0)	5.0 (0.3)	4.4 (0.3)	4.1 (0.3)	4.9 (0.3)
0.75	0.3 (1.0)	7.2 (1.1)	10.7 (1.1)	-0.5 (1.0)	5.0 (0.3)	3.4 (0.3)	3.0 (0.2)	5.3 (0.3)
0.95	1.3 (1.0)	14.8 (1.2)	24.8 (1.3)	1.5 (1.0)	4.7 (0.3)	1.4 (0.2)	0.8 (0.1)	4.7 (0.3)

Table 8.1: Estimated relative error and estimated type I error of linear mixed effects model analysis, adjusting for a covariate when test of balance of covariate is significant at a level of α. 50 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.005. Covariate ICC = 0.005.

	Estimated Relative Error of SE % (MCSE)				Estimated Type I Error % (MCSE)			
Covariate Effect Parameter	$\alpha = 0$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 0$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 1$
0.00	0.5 (1.0)	0.6 (1.0)	0.7 (1.0)	0.7 (1.0)	5.5 (0.3)	5.5 (0.3)	5.4 (0.3)	5.2 (0.3)
0.08	0.3 (1.0)	0.9 (1.0)	1.1 (1.0)	0.0 (1.0)	5.2 (0.3)	5.0 (0.3)	5.0 (0.3)	5.2 (0.3)
0.16	-1.1 (1.0)	0.7 (1.0)	1.9 (1.0)	-1.1 (1.0)	5.8 (0.3)	5.3 (0.3)	5.0 (0.3)	5.4 (0.3)
0.24	-2.6 (1.0)	2.3 (1.0)	5.7 (1.1)	-1.7 (1.0)	5.9 (0.3)	4.5 (0.3)	3.8 (0.3)	5.5 (0.3)
0.30	0.7 (1.0)	9.2 (1.1)	15.5 (1.2)	-0.8 (1.0)	4.7 (0.3)	2.8 (0.2)	1.9 (0.2)	5.5 (0.3)

Table 8.2: Estimated relative error and estimated type I error of linear mixed effects model analysis, adjusting for a covariate when test of balance of covariate is significant at a level of α. 50 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.1. Cluster level covariate..

8.2 Adjusting for covariates used to stratify randomisation

Estimated type I errors and estimated relative errors of model-based standard errors (for a treatment effect of zero) are presented in Tables 8.3 to 8.5 (pages 141 to 143) for unadjusted and adjusted linear mixed effects model analyses, where the covariate has been used to stratify randomisation. The simulated CRTs had 50 clusters in each treatment arm and 30 individuals in each cluster.

Analyses adjusted for the covariate used to stratify randomisation had close to nominal estimated type I error, and estimated relative error up to 6%. Some deviation from nominal type I error and misestimation of standard error is expected due to the small number of clusters, as described in Section 7.4.2.

Analyses using an unadjusted linear mixed effects model produced lower than nominal type I error and overestimated standard error of treatment effect estimates in several scenarios. For a cluster level covariate (Table 8.3, page 141), estimated relative error reached up to 34%, with estimated type I error as low as 0.9%, for an outcome with an ICC of 0.1 (see Figure 8.3, page 139). For a covariate with an ICC of 0.005 (Table 8.4, page 142), estimated relative error reached up to 56%, with estimated type I error as low as 0.1%, for an outcome with an ICC of 0.005 (see Figure 8.4, page 139). For a covariate with an ICC of 0.1 (Table 8.5, page 143), estimated relative error reached up to 55%, with estimated type I error as low as 0.4%, for an outcome with an ICC of 0.1 (see Figure 8.5, 140).

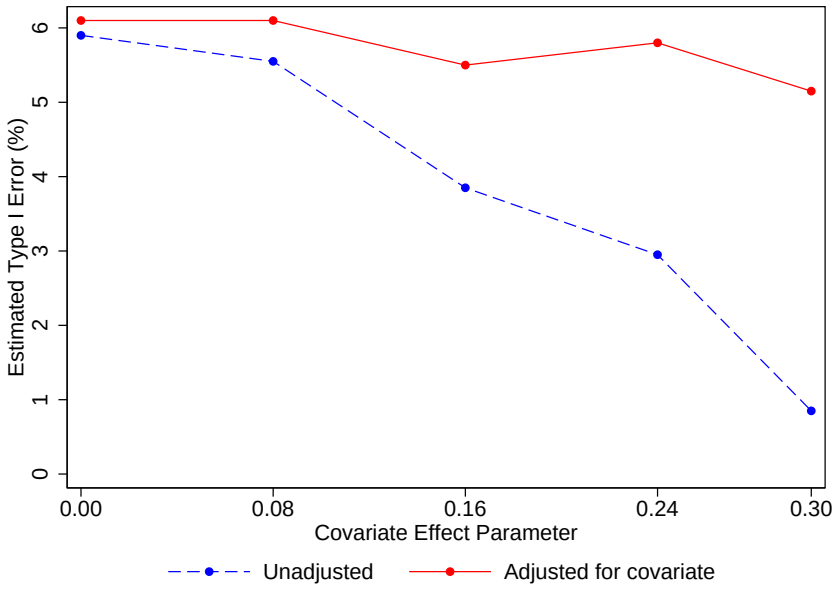


Figure 8.3: Estimated type I error of linear mixed effects model analysis, unadjusted and adjusted for a covariate used to stratify randomisation. 50 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.1. Cluster level covariate. Data from Table 8.3.

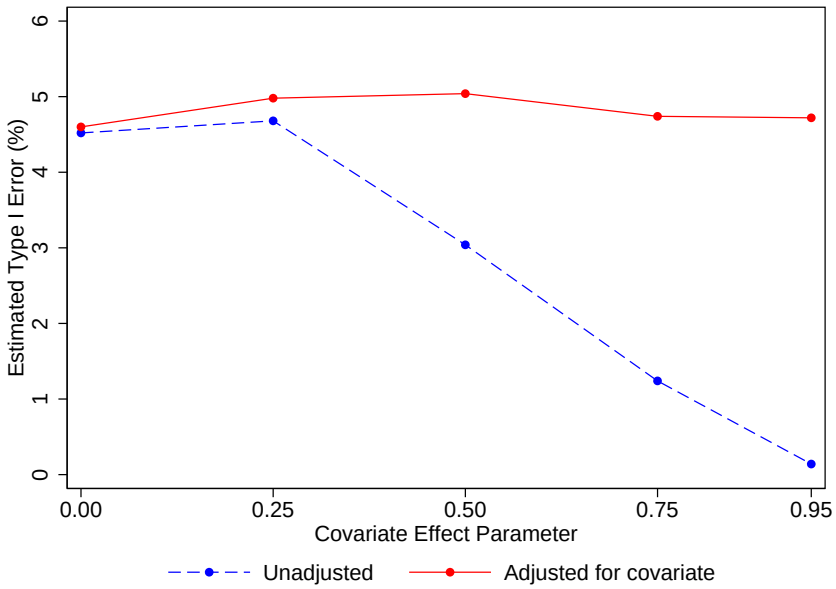


Figure 8.4: Estimated type I error of linear mixed effects model analysis, unadjusted and adjusted for a covariate used to stratify randomisation. 50 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.005. Covariate ICC = 0.005. Data from Table 8.4.

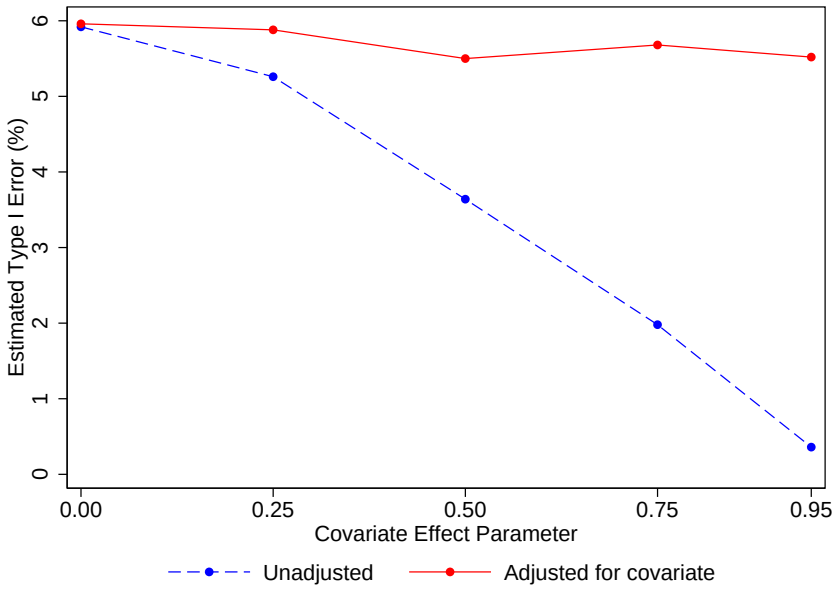


Figure 8.5: Estimated type I error of linear mixed effects model analysis, unadjusted and adjusted for a covariate used to stratify randomisation. 50 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.1. Covariate ICC = 0.1. Data from Table 8.5.

		Estimated Relative Error of SE % (MCSE)		Estimated Type I Error % (MCSE)	
Outcome ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted
0.0005	0.00 (0.000)	6.2 (1.7)	6.3 (1.7)	3.9 (0.4)	3.9 (0.4)
0.0005	0.25 (0.006)	3.5 (1.6)	3.3 (1.6)	4.1 (0.4)	4.3 (0.5)
0.0005	0.50 (0.011)	3.1 (1.6)	3.3 (1.6)	4.5 (0.5)	4.6 (0.5)
0.0005	0.75 (0.017)	5.5 (1.7)	5.6 (1.7)	3.9 (0.4)	3.9 (0.4)
0.0005	0.95 (0.021)	2.4 (1.6)	2.1 (1.6)	4.9 (0.5)	5.1 (0.5)
0.005	0.00 (0.000)	0.8 (1.6)	1.1 (1.6)	5.8 (0.5)	5.6 (0.5)
0.005	0.25 (0.018)	4.1 (1.7)	3.9 (1.7)	4.3 (0.5)	4.4 (0.5)
0.005	0.50 (0.035)	2.3 (1.6)	1.8 (1.6)	4.8 (0.5)	4.6 (0.5)
0.005	0.75 (0.053)	7.6 (1.7)	5.5 (1.7)	3.8 (0.4)	4.0 (0.4)
0.005	0.95 (0.067)	4.1 (1.7)	2.1 (1.6)	4.6 (0.5)	5.0 (0.5)
0.05	0.00 (0.000)	1.7 (1.6)	1.5 (1.6)	5.6 (0.5)	5.5 (0.5)
0.05	0.25 (0.056)	1.4 (1.6)	-0.1 (1.6)	5.6 (0.5)	6.1 (0.5)
0.05	0.50 (0.112)	8.4 (1.7)	3.2 (1.7)	4.0 (0.4)	4.8 (0.5)
0.05	0.75 (0.168)	15.6 (1.9)	2.1 (1.6)	2.9 (0.4)	5.4 (0.5)
0.05	0.95 (0.212)	29.3 (2.1)	3.6 (1.7)	1.7 (0.3)	4.5 (0.5)
0.1	0.00 (0.000)	0.8 (1.6)	1.1 (1.6)	5.9 (0.5)	6.1 (0.5)
0.1	0.25 (0.079)	-0.1 (1.6)	-1.8 (1.6)	5.6 (0.5)	6.1 (0.5)
0.1	0.50 (0.158)	7.2 (1.7)	0.1 (1.6)	3.9 (0.4)	5.5 (0.5)
0.1	0.75 (0.237)	17.7 (1.9)	0.6 (1.6)	3.0 (0.4)	5.8 (0.5)
0.1	0.95 (0.300)	34.4 (2.2)	1.7 (1.6)	0.9 (0.2)	5.2 (0.5)

Table 8.3: Estimated relative error and type I error of linear mixed effects model analysis, unadjusted and adjusted for a covariate used to stratify randomisation. 50 clusters per treatment arm, cluster size of 30. Cluster level covariate.

		Estimated Relative Error of SE % (MCSE)		Estimated Type I Error % (MCSE)	
Outcome ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted
0.0005	0.00 (0.000)	5.6 (1.1)	5.6 (1.1)	4.1 (0.3)	4.2 (0.3)
0.0005	0.25 (0.079)	5.2 (1.1)	4.9 (1.1)	4.1 (0.3)	4.0 (0.3)
0.0005	0.50 (0.158)	3.8 (1.0)	3.1 (1.0)	4.1 (0.3)	4.3 (0.3)
0.0005	0.75 (0.237)	6.2 (1.1)	4.1 (1.0)	4.0 (0.3)	4.4 (0.3)
0.0005	0.95 (0.300)	6.8 (1.1)	3.8 (1.0)	4.0 (0.3)	4.5 (0.3)
0.005	0.00 (0.000)	3.9 (1.0)	3.9 (1.0)	4.5 (0.3)	4.6 (0.3)
0.005	0.25 (0.250)	5.5 (1.1)	3.6 (1.0)	4.7 (0.3)	5.0 (0.3)
0.005	0.50 (0.500)	11.8 (1.1)	2.2 (1.0)	3.0 (0.2)	5.0 (0.3)
0.005	0.75 (0.750)	29.4 (1.3)	3.1 (1.0)	1.2 (0.2)	4.7 (0.3)
0.005	0.95 (0.950)	55.9 (1.6)	2.7 (1.0)	0.1 (0.1)	4.7 (0.3)
0.05	0.00 (0.000)	0.8 (1.0)	0.8 (1.0)	5.7 (0.3)	5.7 (0.3)
0.05	0.25 (0.244)	0.9 (1.0)	0.1 (1.0)	5.6 (0.3)	5.9 (0.3)
0.05	0.50 (0.489)	4.5 (1.1)	0.8 (1.0)	4.6 (0.3)	5.1 (0.3)
0.05	0.75 (0.733)	9.4 (1.1)	0.5 (1.0)	3.9 (0.3)	5.8 (0.3)
0.05	0.95 (0.928)	17.4 (1.2)	0.4 (1.0)	3.0 (0.2)	5.8 (0.3)
0.1	0.00 (0.000)	-0.9 (1.0)	-0.9 (1.0)	6.2 (0.3)	6.2 (0.3)
0.1	0.25 (0.238)	0.2 (1.0)	-0.2 (1.0)	5.5 (0.3)	5.6 (0.3)
0.1	0.50 (0.476)	2.2 (1.0)	-0.2 (1.0)	5.1 (0.3)	5.4 (0.3)
0.1	0.75 (0.713)	5.2 (1.1)	-0.5 (1.0)	4.6 (0.3)	5.9 (0.3)
0.1	0.95 (0.904)	7.8 (1.1)	-1.1 (1.0)	4.4 (0.3)	6.3 (0.3)

Table 8.4: Estimated relative error and type I error of linear mixed effects model analysis, unadjusted and adjusted for a covariate used to stratify randomisation. 50 clusters per treatment arm, cluster size of 30. Covariate ICC = 0.005.

		Estimated Relative Error of SE % (MCSE)		Estimated Type I Error % (MCSE)	
Outcome ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted
0.0005	0.00 (0.000)	3.5 (1.0)	3.6 (1.0)	4.5 (0.3)	4.4 (0.3)
0.0005	0.25 (0.018)	4.9 (1.1)	5.0 (1.1)	4.4 (0.3)	4.5 (0.3)
0.0005	0.50 (0.035)	4.9 (1.1)	4.9 (1.1)	4.1 (0.3)	4.1 (0.3)
0.0005	0.75 (0.053)	4.6 (1.1)	4.5 (1.1)	4.4 (0.3)	4.4 (0.3)
0.0005	0.95 (0.067)	3.1 (1.0)	2.8 (1.0)	4.6 (0.3)	4.8 (0.3)
0.005	0.00 (0.000)	1.7 (1.0)	1.7 (1.0)	5.4 (0.3)	5.5 (0.3)
0.005	0.25 (0.056)	2.6 (1.0)	2.3 (1.0)	4.9 (0.3)	5.0 (0.3)
0.005	0.50 (0.112)	3.6 (1.0)	2.3 (1.0)	4.4 (0.3)	4.7 (0.3)
0.005	0.75 (0.168)	6.3 (1.1)	4.0 (1.0)	3.8 (0.3)	4.6 (0.3)
0.005	0.95 (0.212)	7.0 (1.1)	3.9 (1.0)	4.0 (0.3)	4.8 (0.3)
0.05	0.00 (0.000)	1.4 (1.0)	1.4 (1.0)	5.6 (0.3)	5.5 (0.3)
0.05	0.25 (0.177)	2.0 (1.0)	0.0 (1.0)	5.1 (0.3)	5.7 (0.3)
0.05	0.50 (0.354)	7.3 (1.1)	0.3 (1.0)	4.1 (0.3)	5.4 (0.3)
0.05	0.75 (0.530)	18.2 (1.2)	-0.4 (1.0)	2.8 (0.2)	5.7 (0.3)
0.05	0.95 (0.672)	36.8 (1.4)	-0.4 (1.0)	1.1 (0.1)	5.4 (0.3)
0.1	0.00 (0.000)	-0.7 (1.0)	-0.8 (1.0)	5.9 (0.3)	6.0 (0.3)
0.1	0.25 (0.250)	2.7 (1.0)	1.0 (1.0)	5.3 (0.3)	5.9 (0.3)
0.1	0.50 (0.500)	10.7 (1.1)	0.9 (1.0)	3.6 (0.3)	5.5 (0.3)
0.1	0.75 (0.750)	22.8 (1.2)	-0.1 (1.0)	2.0 (0.2)	5.7 (0.3)
0.1	0.95 (0.950)	55.0 (1.6)	0.7 (1.0)	0.4 (0.1)	5.5 (0.3)

Table 8.5: Estimated relative error and type I error of linear mixed effects model analysis, unadjusted and adjusted for a covariate used to stratify randomisation. 50 clusters per treatment arm, cluster size of 30. Covariate ICC = 0.1.

8.3 Effects of covariate adjustment in linear mixed effects models

In this section I present empirical standard errors and estimated power from simulations of unadjusted and adjusted linear mixed effects model analyses.

In all situations, adjusting for a covariate with no effect on the outcome did not significantly change the estimated power of the analysis or the standard error of estimated treatment effect. As covariate effect increased, estimated power of adjusted analyses increased and empirical standard errors from adjusted analyses became smaller.

Empirical standard errors and estimated power from simulations of unadjusted and adjusted linear mixed effects model analyses are presented in Tables 8.6 to 8.9, for simulated CRTs with a cluster size of 30 and 20 clusters in each treatment arm. The tables also present the power of the adjusted analyses as calculated using equation 7.21.

Results for an outcome with an ICC of 0.0005 are presented in Table 8.6 (page 146). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 100% from 78% (see Figure 8.6, page 145), reducing standard error from 0.058 to 0.018. Adjusting for a covariate with an ICC of 0.05 did not significantly increase estimated power or reduce empirical standard error for any size of covariate effect. Adjusting for a cluster level covariate also did not increase estimated power or reduce empirical standard error for any size of covariate effect.

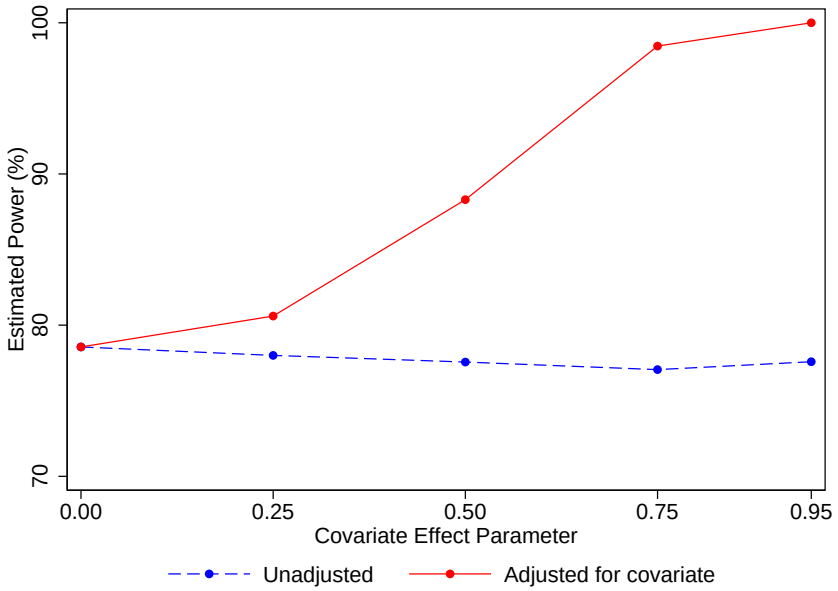


Figure 8.6: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.0005. Covariate ICC = 0.0005. Treatment effect parameter = 0.163. Data from Table 8.6.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.059 (0.001)	0.059 (0.001)	78.6 (0.6)	78.6 (0.6)	77.9
0.0005	0.25 (0.250)	0.058 (0.001)	0.056 (0.001)	78.0 (0.6)	80.6 (0.6)	80.4
0.0005	0.50 (0.500)	0.059 (0.001)	0.051 (0.001)	77.6 (0.6)	88.3 (0.5)	88.3
0.0005	0.75 (0.750)	0.059 (0.001)	0.039 (<0.001)	77.1 (0.6)	98.5 (0.2)	98.5
0.0005	0.95 (0.950)	0.058 (0.001)	0.018 (<0.001)	77.6 (0.6)	100.0 -	100.0
0.1	0.00 (0.000)	0.058 (0.001)	0.058 (0.001)	76.8 (0.6)	76.8 (0.6)	77.9
0.1	0.25 (0.018)	0.058 (0.001)	0.058 (0.001)	77.8 (0.6)	77.7 (0.6)	77.9
0.1	0.50 (0.035)	0.058 (0.001)	0.058 (0.001)	77.7 (0.6)	77.7 (0.6)	78.1
0.1	0.75 (0.053)	0.058 (0.001)	0.057 (0.001)	77.5 (0.6)	77.2 (0.6)	78.3
0.1	0.95 (0.067)	0.059 (0.001)	0.059 (0.001)	76.6 (0.6)	77.0 (0.6)	78.6
1	0.00 (0.000)	0.058 (0.001)	0.059 (0.001)	77.6 (0.6)	76.4 (0.6)	77.9
1	0.25 (0.006)	0.058 (0.001)	0.059 (0.001)	78.6 (0.6)	77.2 (0.6)	77.9
1	0.50 (0.011)	0.058 (0.001)	0.059 (0.001)	77.9 (0.6)	76.6 (0.6)	78.0
1	0.75 (0.017)	0.058 (0.001)	0.059 (0.001)	77.5 (0.6)	77.2 (0.6)	78.2
1	0.95 (0.021)	0.058 (0.001)	0.058 (0.001)	77.4 (0.6)	77.2 (0.6)	78.4

Table 8.6: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.0005. Treatment effect parameter = 0.163. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.005 are presented in Table 8.7 (page 149). Adjusting for a covariate with an ICC of 0.0005 or an ICC of 0.005 increased estimated power to a maximum of 100% (see Figures 8.7 and 8.8, page 148). Adjusting for a covariate with an ICC of 0.0005 reduced empirical standard error from 0.062 to 0.028, and adjusting for a covariate with an ICC of 0.005 reduced empirical standard error from 0.061 to 0.019. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 84% from 79% (see Figure 8.9, page 148), reducing empirical standard error from 0.063 to 0.058.

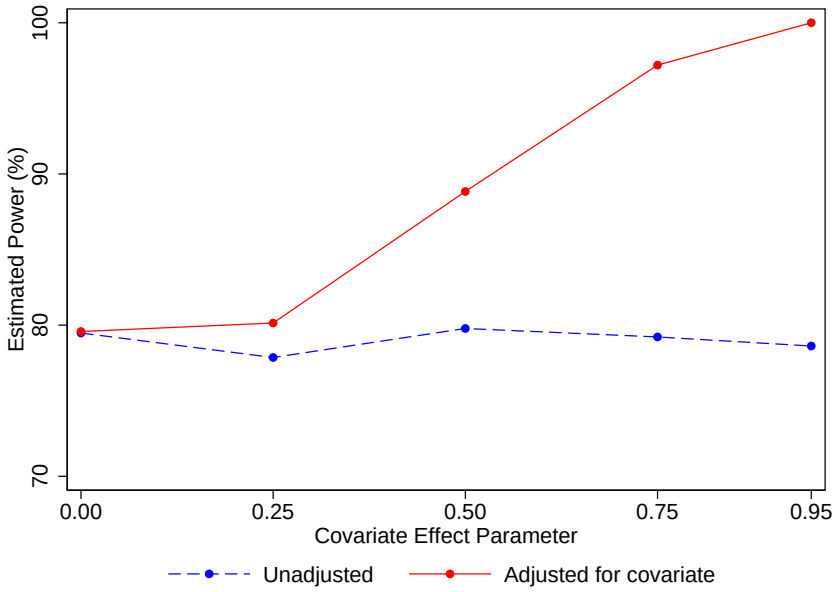


Figure 8.7: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.005. Covariate ICC = 0.0005. Treatment effect parameter = 0.173. Data from Table 8.7.

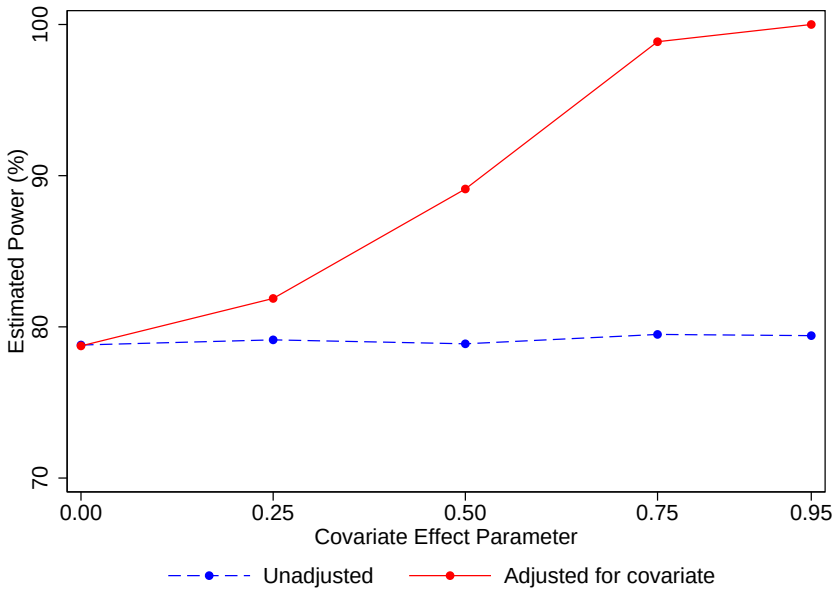


Figure 8.8: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.005. Covariate ICC = 0.005. Treatment effect parameter = 0.173. Data from Table 8.7.

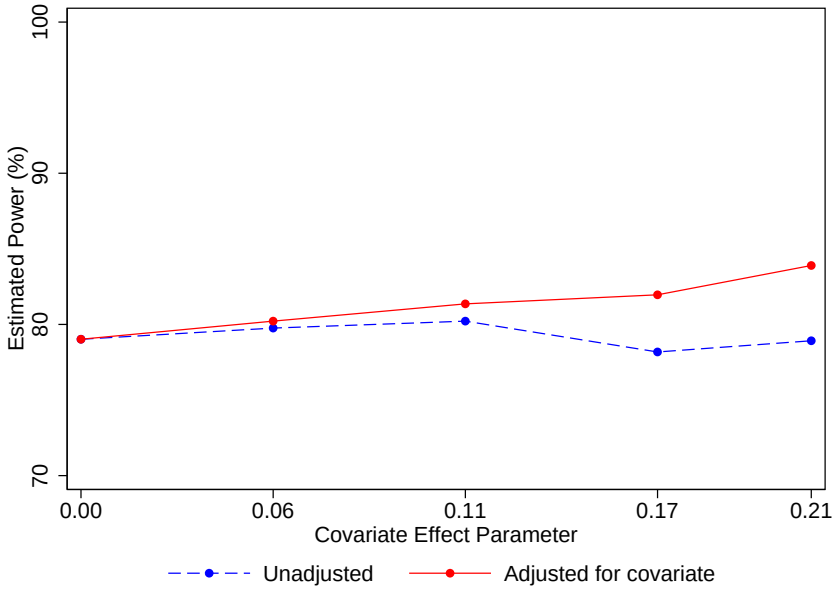


Figure 8.9: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.005. Covariate ICC = 0.1. Treatment effect parameter = 0.173. Data from Table 8.7.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.060 (0.001)	0.060 (0.001)	79.5 (0.6)	79.6 (0.6)	77.9
0.0005	0.25 (0.249)	0.063 (0.001)	0.061 (0.001)	77.9 (0.6)	80.1 (0.6)	80.1
0.0005	0.50 (0.499)	0.062 (0.001)	0.054 (0.001)	79.8 (0.6)	88.8 (0.4)	87.1
0.0005	0.75 (0.748)	0.061 (0.001)	0.044 (<0.001)	79.2 (0.6)	97.2 (0.2)	97.0
0.0005	0.95 (0.948)	0.062 (0.001)	0.028 (<0.001)	78.6 (0.6)	100.0 -	100.0
0.005	0.00 (0.000)	0.063 (0.001)	0.063 (0.001)	78.8 (0.6)	78.7 (0.6)	77.9
0.005	0.25 (0.250)	0.062 (0.001)	0.060 (0.001)	79.1 (0.6)	81.9 (0.5)	80.4
0.005	0.50 (0.500)	0.062 (0.001)	0.053 (0.001)	78.9 (0.6)	89.1 (0.4)	88.3
0.005	0.75 (0.750)	0.061 (0.001)	0.041 (<0.001)	79.5 (0.6)	98.9 (0.2)	98.5
0.005	0.95 (0.950)	0.061 (0.001)	0.019 (<0.001)	79.4 (0.6)	100.0 -	100.0

Continued on next page.

Table 8.7: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 20 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.005. Treatment effect parameter = 0.173. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.1	0.00 (0.000)	0.061 (0.001)	0.061 (0.001)	79.0 (0.6)	79.0 (0.6)	77.9
0.1	0.25 (0.056)	0.061 (0.001)	0.060 (0.001)	79.8 (0.6)	80.2 (0.6)	78.3
0.1	0.50 (0.112)	0.061 (0.001)	0.060 (0.001)	80.2 (0.6)	81.4 (0.6)	79.6
0.1	0.75 (0.168)	0.061 (0.001)	0.058 (0.001)	78.2 (0.6)	82.0 (0.5)	81.8
0.1	0.95 (0.212)	0.063 (0.001)	0.058 (0.001)	78.9 (0.6)	83.9 (0.5)	84.3
1	0.00 (0.000)	0.062 (0.001)	0.063 (0.001)	79.0 (0.6)	78.2 (0.6)	77.9
1	0.25 (0.018)	0.062 (0.001)	0.063 (0.001)	78.6 (0.6)	77.7 (0.6)	78.2
1	0.50 (0.035)	0.062 (0.001)	0.062 (0.001)	78.0 (0.6)	77.8 (0.6)	79.2
1	0.75 (0.053)	0.062 (0.001)	0.061 (0.001)	77.4 (0.6)	79.5 (0.6)	80.9
1	0.95 (0.067)	0.061 (0.001)	0.058 (0.001)	79.1 (0.6)	82.0 (0.5)	82.8

Table 8.7: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 20 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.005. Treatment effect parameter = 0.173. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.05 are presented in Table 8.8 (page 153). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 93% from 81% (see Figure 8.10). Adjusting for a covariate with an ICC of 0.05 increased estimated power to a maximum of 100% from 80% (see Figure 8.11, page 152), reducing empirical standard error from 0.091 to 0.028. Adjusting for a cluster level covariate increased estimated power to a maximum of 98% from 80% (see Figure 8.12, page 152), reducing empirical standard error from 0.090 to 0.060.

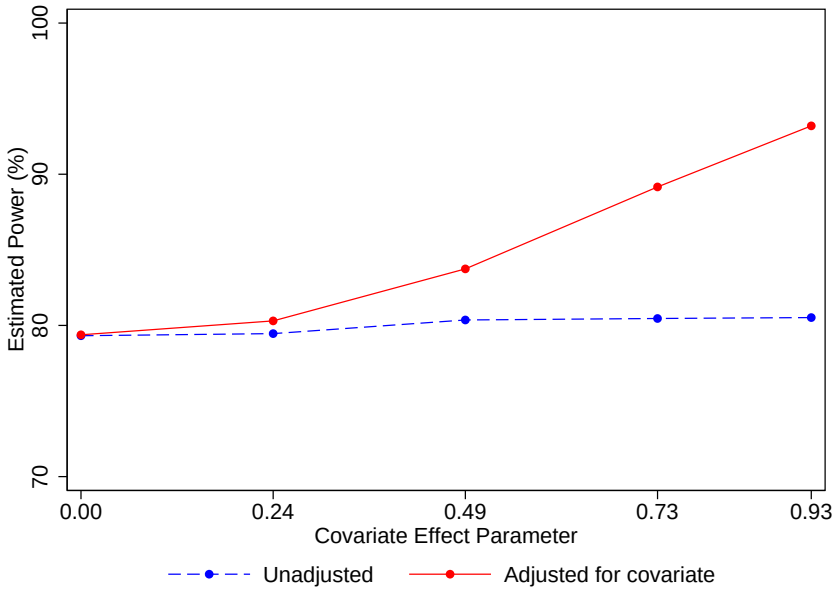


Figure 8.10: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Covariate ICC = 0.0005. Treatment effect parameter = 0.253. Data from Table 8.8.

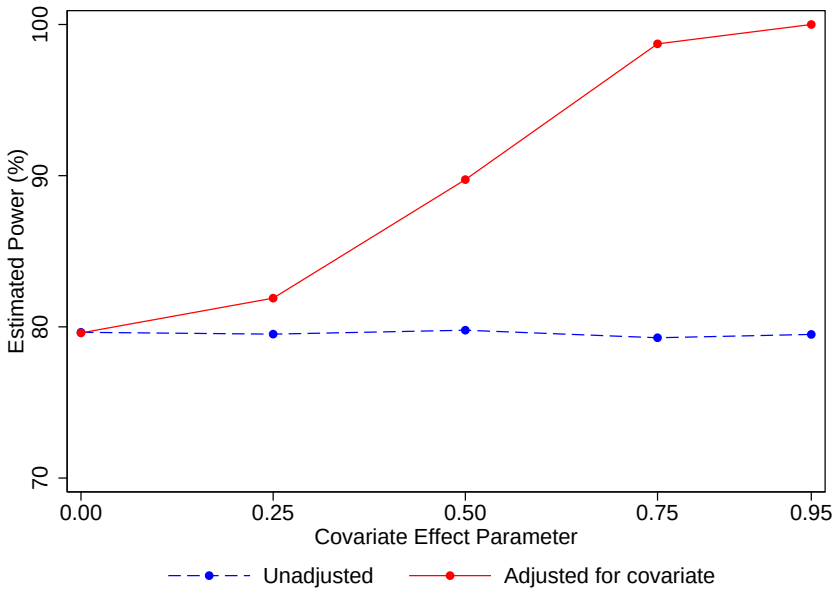


Figure 8.11: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Covariate ICC = 0.05. Treatment effect parameter = 0.253. Data from Table 8.8.

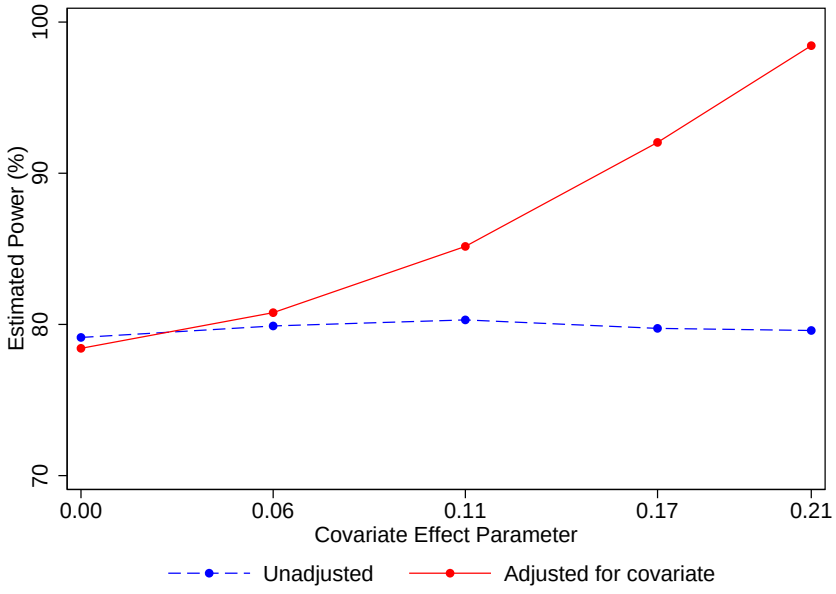


Figure 8.12: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Covariate ICC = 0.1. Treatment effect parameter = 0.253. Data from Table 8.8.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.093 (0.001)	0.093 (0.001)	79.3 (0.6)	79.4 (0.6)	77.9
0.0005	0.25 (0.244)	0.091 (0.001)	0.089 (0.001)	79.5 (0.6)	80.3 (0.6)	78.9
0.0005	0.50 (0.487)	0.089 (0.001)	0.085 (0.001)	80.4 (0.6)	83.7 (0.5)	81.9
0.0005	0.75 (0.731)	0.090 (0.001)	0.079 (0.001)	80.5 (0.6)	89.2 (0.4)	87.1
0.0005	0.95 (0.926)	0.091 (0.001)	0.073 (0.001)	80.5 (0.6)	93.2 (0.4)	92.5
0.05	0.00 (0.000)	0.092 (0.001)	0.092 (0.001)	79.6 (0.6)	79.6 (0.6)	77.9
0.05	0.25 (0.250)	0.091 (0.001)	0.089 (0.001)	79.5 (0.6)	81.9 (0.5)	80.4
0.05	0.50 (0.500)	0.091 (0.001)	0.079 (0.001)	79.8 (0.6)	89.7 (0.4)	88.3
0.05	0.75 (0.750)	0.090 (0.001)	0.059 (0.001)	79.3 (0.6)	98.7 (0.2)	98.5
0.05	0.95 (0.950)	0.091 (0.001)	0.028 (<0.001)	79.5 (0.6)	100.0 -	100.0

Continued on next page.

Table 8.8: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 20 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.05. Treatment effect parameter = 0.253. (CEF=Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.1	0.00 (0.000)	0.091 (0.001)	0.091 (0.001)	79.1 (0.6)	79.1 (0.6)	77.9
0.1	0.25 (0.177)	0.090 (0.001)	0.087 (0.001)	78.9 (0.6)	81.4 (0.5)	79.9
0.1	0.50 (0.354)	0.089 (0.001)	0.080 (0.001)	80.0 (0.6)	88.1 (0.5)	86.2
0.1	0.75 (0.530)	0.090 (0.001)	0.067 (0.001)	79.0 (0.6)	96.1 (0.3)	95.6
0.1	0.95 (0.672)	0.090 (0.001)	0.049 (<0.001)	79.5 (0.6)	99.9 (<0.1)	99.9
1	0.00 (0.000)	0.091 (0.001)	0.092 (0.001)	79.1 (0.6)	78.4 (0.6)	77.9
1	0.25 (0.056)	0.090 (0.001)	0.090 (0.001)	79.9 (0.6)	80.8 (0.6)	79.4
1	0.50 (0.112)	0.089 (0.001)	0.084 (0.001)	80.3 (0.6)	85.2 (0.5)	84.2
1	0.75 (0.168)	0.090 (0.001)	0.074 (0.001)	79.7 (0.6)	92.0 (0.4)	92.1
1	0.95 (0.212)	0.090 (0.001)	0.060 (0.001)	79.6 (0.6)	98.4 (0.2)	98.3

Table 8.8: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 20 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.05. Treatment effect parameter = 0.253. (CEF=Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.1 are presented in Table 8.9 (page 157). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 88% from 80% (see Figure 8.13). Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 100% from 80% (see Figure 8.14, page 156), reducing empirical standard error from 0.113 to 0.036. Adjusting for a cluster level covariate increased estimated power to a maximum of 99.9% from 79% (see Figure 8.15, page 156).

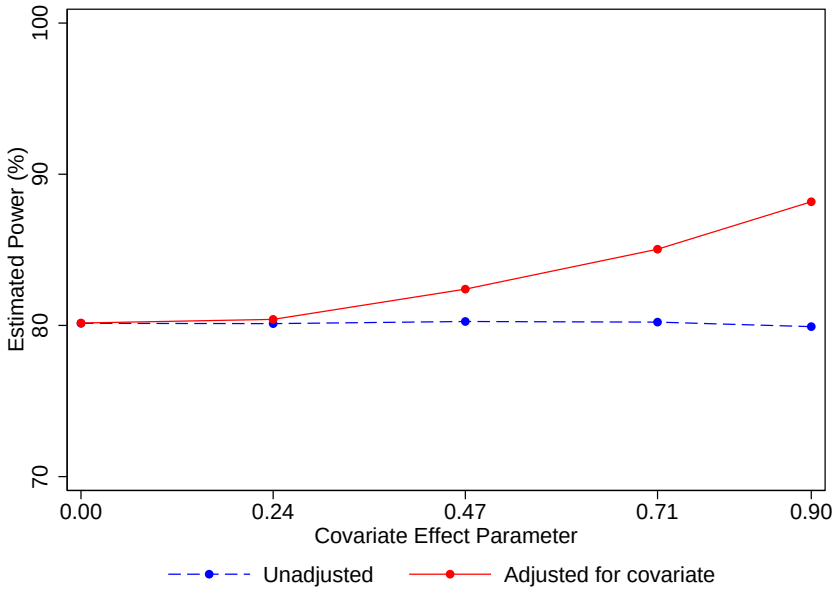


Figure 8.13: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. Covariate ICC = 0.0005. Treatment effect parameter = 0.31943. Data from Table 8.9.

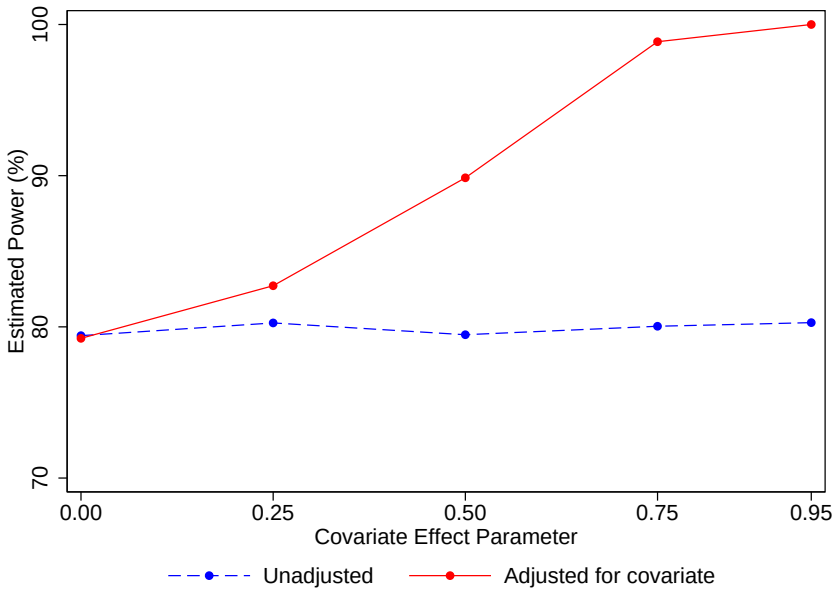


Figure 8.14: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. Covariate ICC = 0.1. Treatment effect parameter = 0.319. Data from Table 8.9.

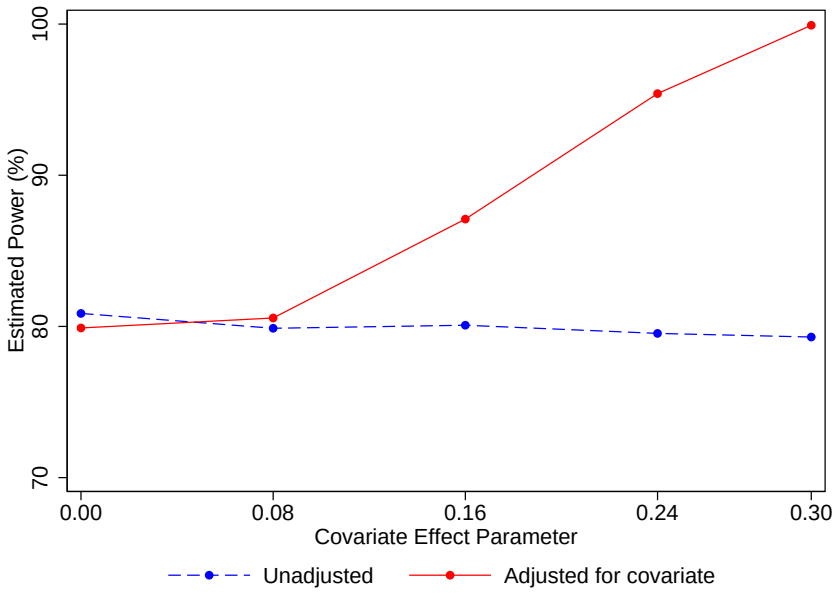


Figure 8.15: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. Cluster level covariate. Treatment effect parameter = 0.319. Data from Table 8.9.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.113 (0.001)	0.113 (0.001)	80.1 (0.6)	80.2 (0.6)	77.9
0.0005	0.25 (0.237)	0.113 (0.001)	0.112 (0.001)	80.1 (0.6)	80.4 (0.6)	78.5
0.0005	0.50 (0.474)	0.112 (0.001)	0.109 (0.001)	80.3 (0.6)	82.4 (0.5)	80.3
0.0005	0.75 (0.712)	0.114 (0.001)	0.107 (0.001)	80.2 (0.6)	85.0 (0.5)	83.3
0.0005	0.95 (0.901)	0.113 (0.001)	0.101 (0.001)	79.9 (0.6)	88.2 (0.5)	86.7
0.1	0.00 (0.000)	0.112 (0.001)	0.112 (0.001)	79.4 (0.6)	79.2 (0.6)	77.9
0.1	0.25 (0.250)	0.113 (0.001)	0.110 (0.001)	80.3 (0.6)	82.7 (0.5)	80.4
0.1	0.50 (0.500)	0.115 (0.001)	0.098 (0.001)	79.5 (0.6)	89.9 (0.4)	88.3
0.1	0.75 (0.750)	0.115 (0.001)	0.076 (0.001)	80.0 (0.6)	98.9 (0.2)	98.5
0.1	0.95 (0.950)	0.113 (0.001)	0.036 (<0.001)	80.3 (0.6)	100.0 -	100.0
1	0.00 (0.000)	0.113 (0.001)	0.115 (0.001)	80.9 (0.6)	79.9 (0.6)	77.9
1	0.25 (0.079)	0.115 (0.001)	0.113 (0.001)	79.9 (0.6)	80.6 (0.6)	79.8
1	0.50 (0.158)	0.114 (0.001)	0.103 (0.001)	80.1 (0.6)	87.1 (0.5)	85.9
1	0.75 (0.237)	0.114 (0.001)	0.087 (0.001)	79.5 (0.6)	95.4 (0.3)	95.2
1	0.95 (0.300)	0.114 (0.001)	0.064 (0.001)	79.3 (0.6)	99.9 (<0.1)	99.9

Table 8.9: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 20 clusters per treatment arm, cluster size of 30. Outcome ICC = 0.1. Treatment effect parameter = 0.319. (CEF=Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Tables 8.10 to 8.13 present empirical standard error and estimated power under unadjusted and adjusted analyses for simulated CRTs with 60 clusters in each treatment arm and a cluster size of five.

Results for an outcome with an ICC of 0.0005 are presented in Table 8.10 (page 159). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 100% from 79% (see Figure 8.16), reducing empirical standard error from 0.082 to 0.026. Adjusting for a covariate with an ICC of 0.05 or a cluster level covariate did not significantly increase estimated power or reduce empirical standard error. Power calculated from the formula suggest that these would increase power by less than 1%.

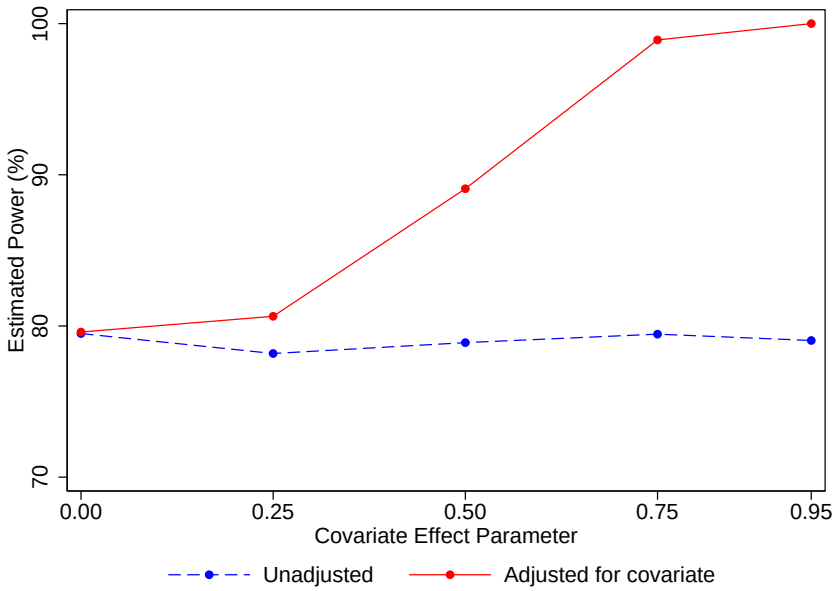


Figure 8.16: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.0005. Covariate ICC = 0.0005. Treatment effect parameter = 0.229. Data from Table 8.10.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.081 (0.001)	0.081 (0.001)	79.5 (0.6)	79.6 (0.6)	79.3
0.0005	0.25 (0.250)	0.082 (0.001)	0.080 (0.001)	78.2 (0.6)	80.6 (0.6)	81.9
0.0005	0.50 (0.500)	0.081 (0.001)	0.070 (0.001)	78.9 (0.6)	89.1 (0.4)	89.4
0.0005	0.75 (0.750)	0.083 (0.001)	0.054 (0.001)	79.5 (0.6)	98.9 (0.1)	98.7
0.0005	0.95 (0.950)	0.082 (0.001)	0.026 (<0.001)	79.0 (0.6)	100.0 -	100.0
0.1	0.00 (0.000)	0.083 (0.001)	0.084 (0.001)	78.4 (0.6)	78.3 (0.6)	79.3
0.1	0.25 (0.018)	0.081 (0.001)	0.081 (0.001)	78.7 (0.6)	78.7 (0.6)	79.4
0.1	0.50 (0.035)	0.082 (0.001)	0.082 (0.001)	78.4 (0.6)	78.5 (0.6)	79.4
0.1	0.75 (0.053)	0.081 (0.001)	0.081 (0.001)	79.5 (0.6)	79.4 (0.6)	79.5
0.1	0.95 (0.067)	0.081 (0.001)	0.081 (0.001)	79.1 (0.6)	79.2 (0.6)	79.6
1	0.00 (0.000)	0.083 (0.001)	0.083 (0.001)	79.4 (0.6)	79.3 (0.6)	79.3
1	0.25 (0.006)	0.081 (0.001)	0.082 (0.001)	78.8 (0.6)	78.1 (0.6)	79.4
1	0.50 (0.011)	0.083 (0.001)	0.083 (0.001)	77.8 (0.6)	77.9 (0.6)	79.4
1	0.75 (0.017)	0.081 (0.001)	0.081 (0.001)	78.2 (0.6)	77.5 (0.6)	79.4
1	0.95 (0.021)	0.082 (0.001)	0.083 (0.001)	79.3 (0.6)	79.0 (0.6)	79.4

Table 8.10: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 60 clusters per treatment arm, cluster size of 5. Outcome ICC = 0.0005. Treatment effect parameter = 0.229. (CEF=Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.005 are presented in Table 8.11 (page 162). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 100% from 80% (see Figure 8.17), reducing empirical standard error from 0.082 to 0.029. Adjusting for a covariate with an ICC of 0.005 increased estimated power to a maximum of 100% from 79% (see Figure 8.18, page 161), reducing standard error from 0.081 to 0.026. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 83% (see Figure 8.19, page 161).

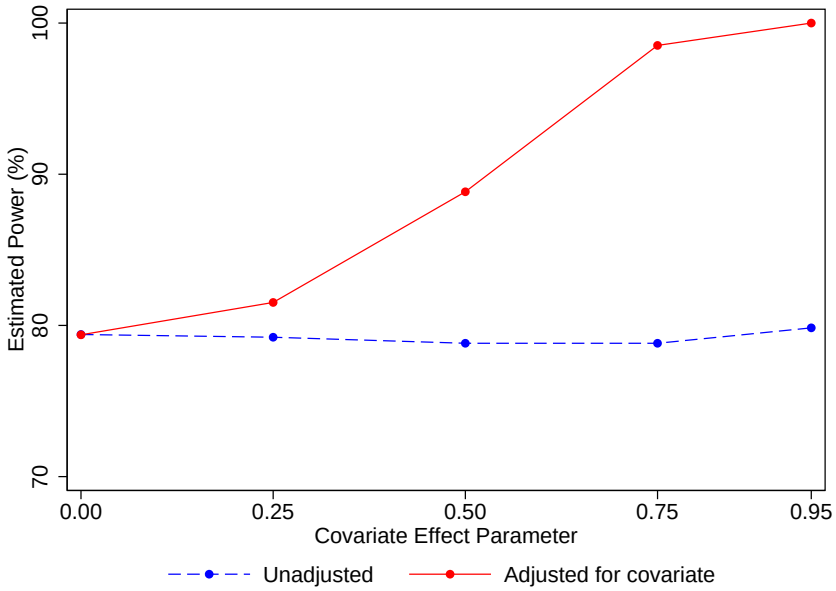


Figure 8.17: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.005. Covariate ICC = 0.0005. Treatment effect parameter = 0.231. Data from Table 8.11.

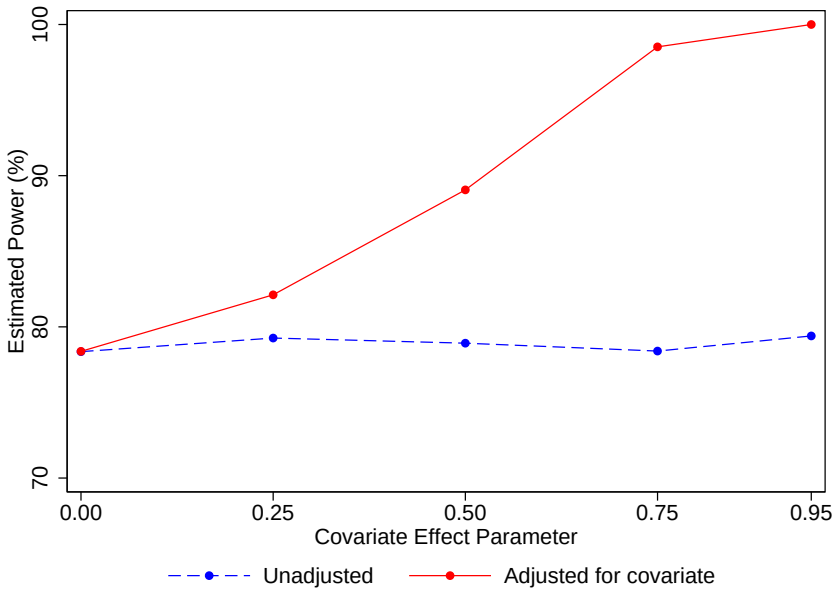


Figure 8.18: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.005. Covariate ICC = 0.005. Treatment effect parameter = 0.231. Data from Table 8.11.

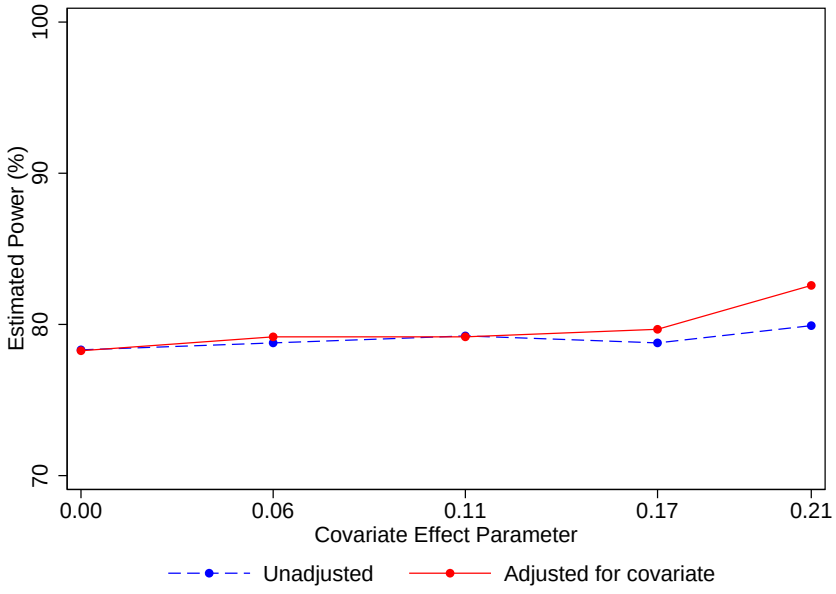


Figure 8.19: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.005. Covariate ICC = 0.1. Treatment effect parameter = 0.231. Data from Table 8.11.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.083 (0.001)	0.083 (0.001)	79.4 (0.6)	79.4 (0.6)	79.3
0.0005	0.25 (0.249)	0.082 (0.001)	0.080 (0.001)	79.2 (0.6)	81.5 (0.5)	81.8
0.0005	0.50 (0.499)	0.083 (0.001)	0.071 (0.001)	78.8 (0.6)	88.8 (0.4)	89.2
0.0005	0.75 (0.748)	0.084 (0.001)	0.056 (0.001)	78.8 (0.6)	98.5 (0.2)	98.5
0.0005	0.95 (0.948)	0.082 (0.001)	0.029 (<0.001)	79.8 (0.6)	100.0 -	100.0
0.005	0.00 (0.000)	0.082 (0.001)	0.082 (0.001)	78.4 (0.6)	78.4 (0.6)	79.3
0.005	0.25 (0.250)	0.082 (0.001)	0.080 (0.001)	79.3 (0.6)	82.1 (0.5)	81.9
0.005	0.50 (0.500)	0.082 (0.001)	0.071 (0.001)	78.9 (0.6)	89.1 (0.4)	89.4
0.005	0.75 (0.750)	0.082 (0.001)	0.054 (0.001)	78.4 (0.6)	98.5 (0.2)	98.7
0.005	0.95 (0.950)	0.081 (0.001)	0.026 (<0.001)	79.4 (0.6)	100.0 -	100.0

Continued on next page.

Table 8.11: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 60 clusters per treatment arm, cluster size of 5. Outcome ICC = 0.005. Treatment effect parameter = 0.231. (CEF=Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.1	0.00 (0.000)	0.083 (0.001)	0.083 (0.001)	78.3 (0.6)	78.3 (0.6)	79.3
0.1	0.25 (0.056)	0.083 (0.001)	0.083 (0.001)	78.8 (0.6)	79.2 (0.6)	79.5
0.1	0.50 (0.112)	0.083 (0.001)	0.083 (0.001)	79.2 (0.6)	79.2 (0.6)	80.0
0.1	0.75 (0.168)	0.082 (0.001)	0.080 (0.001)	78.8 (0.6)	79.7 (0.6)	80.9
0.1	0.95 (0.212)	0.083 (0.001)	0.080 (0.001)	79.9 (0.6)	82.6 (0.5)	81.8
1	0.00 (0.000)	0.081 (0.001)	0.082 (0.001)	78.6 (0.6)	78.4 (0.6)	79.3
1	0.25 (0.018)	0.084 (0.001)	0.084 (0.001)	78.2 (0.6)	77.6 (0.6)	79.4
1	0.50 (0.035)	0.083 (0.001)	0.083 (0.001)	78.3 (0.6)	78.2 (0.6)	79.6
1	0.75 (0.053)	0.083 (0.001)	0.082 (0.001)	78.4 (0.6)	78.7 (0.6)	79.9
1	0.95 (0.067)	0.083 (0.001)	0.082 (0.001)	79.1 (0.6)	79.4 (0.6)	80.2

Table 8.11: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 60 clusters per treatment arm, cluster size of 5. Outcome ICC = 0.005. Treatment effect parameter = 0.231. (CEF=Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.05 are presented in table 8.12 (page 166). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 100% from 79% (see Figure 8.20), reducing empirical standard error from 0.091 to 0.047. Adjusting for a covariate with an ICC of 0.05 increased estimated power to a maximum of 100% from 81%, reducing empirical standard error from 0.090 to 0.028. Adjusting for a cluster level covariate increased estimated power to a maximum of 86% (see Figure 8.21, page 165), reducing empirical standard error from 0.091 to 0.082.

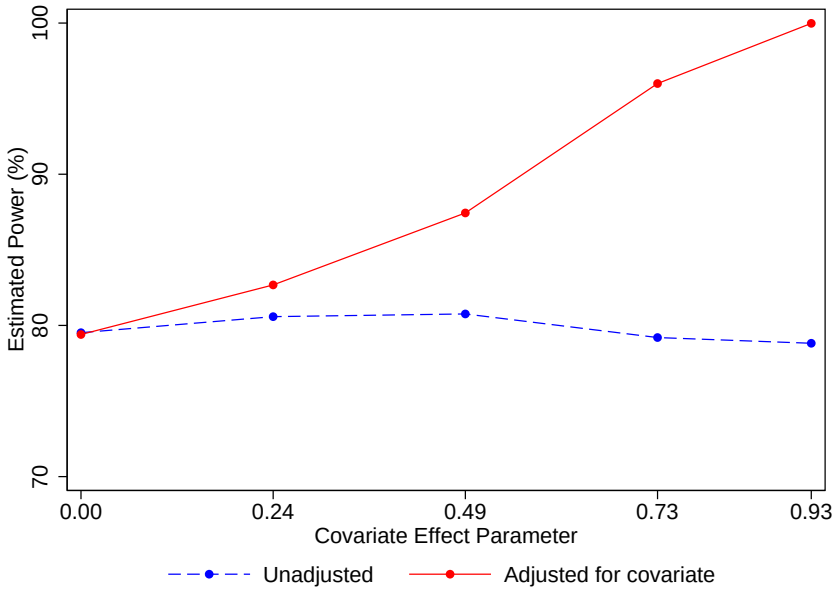


Figure 8.20: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Covariate ICC = 0.0005. Treatment effect parameter = 0.251. Data from Table 8.12.

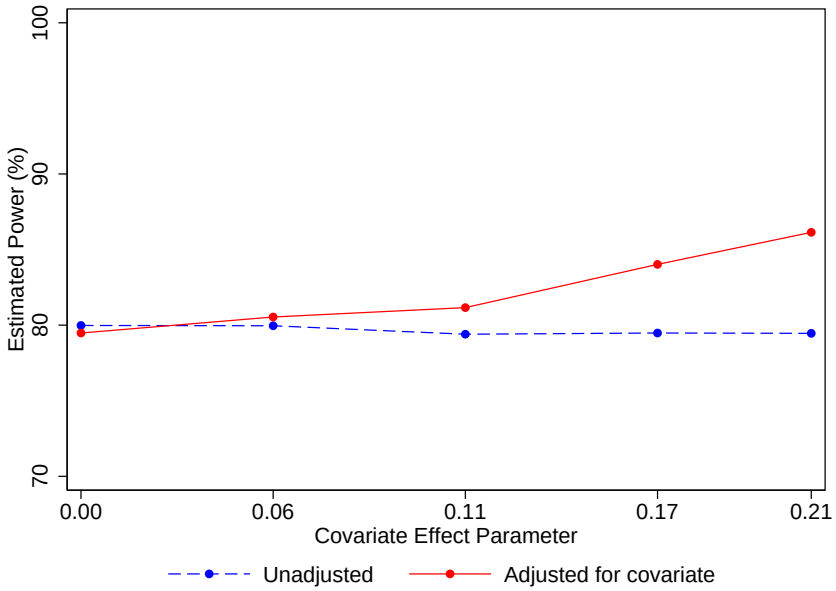


Figure 8.21: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Cluster level covariate. Treatment effect parameter = 0.251. Data from Table 8.12.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.090 (0.001)	0.090 (0.001)	79.5 (0.6)	79.4 (0.6)	79.3
0.0005	0.25 (0.244)	0.090 (0.001)	0.087 (0.001)	80.6 (0.6)	82.7 (0.5)	81.3
0.0005	0.50 (0.487)	0.091 (0.001)	0.081 (0.001)	80.8 (0.6)	87.4 (0.5)	87.4
0.0005	0.75 (0.731)	0.090 (0.001)	0.067 (0.001)	79.2 (0.6)	96.0 (0.3)	96.2
0.0005	0.95 (0.926)	0.091 (0.001)	0.047 (<0.001)	78.8 (0.6)	100.0 -	99.9
0.05	0.00 (0.000)	0.089 (0.001)	0.090 (0.001)	79.6 (0.6)	79.4 (0.6)	79.3
0.05	0.25 (0.250)	0.089 (0.001)	0.086 (0.001)	79.7 (0.6)	82.1 (0.5)	81.9
0.05	0.50 (0.500)	0.088 (0.001)	0.077 (0.001)	81.4 (0.6)	90.4 (0.4)	89.4
0.05	0.75 (0.750)	0.089 (0.001)	0.059 (0.001)	79.9 (0.6)	99.0 (0.1)	98.7
0.05	0.95 (0.950)	0.090 (0.001)	0.028 (<0.001)	80.6 (0.6)	100.0 -	100.0

Continued on next page.

Table 8.12: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 60 clusters per treatment arm, cluster size of 5. Outcome ICC = 0.05. Treatment effect parameter = 0.251. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.1	0.00 (0.000)	0.090 (0.001)	0.090 (0.001)	79.7 (0.6)	79.5 (0.6)	79.3
0.1	0.25 (0.177)	0.089 (0.001)	0.088 (0.001)	80.6 (0.6)	81.5 (0.5)	80.8
0.1	0.50 (0.354)	0.089 (0.001)	0.083 (0.001)	78.8 (0.6)	85.1 (0.5)	85.2
0.1	0.75 (0.530)	0.091 (0.001)	0.074 (0.001)	79.9 (0.6)	92.4 (0.4)	92.4
0.1	0.95 (0.672)	0.089 (0.001)	0.061 (0.001)	80.0 (0.6)	98.1 (0.2)	98.1
1	0.00 (0.000)	0.090 (0.001)	0.091 (0.001)	80.0 (0.6)	79.5 (0.6)	79.3
1	0.25 (0.056)	0.090 (0.001)	0.090 (0.001)	80.0 (0.6)	80.5 (0.6)	79.9
1	0.50 (0.112)	0.089 (0.001)	0.087 (0.001)	79.4 (0.6)	81.2 (0.6)	81.4
1	0.75 (0.168)	0.088 (0.001)	0.084 (0.001)	79.5 (0.6)	84.0 (0.5)	84.1
1	0.95 (0.212)	0.091 (0.001)	0.082 (0.001)	79.5 (0.6)	86.1 (0.5)	86.9

Table 8.12: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 60 clusters per treatment arm, cluster size of 5. Outcome ICC = 0.05. Treatment effect parameter = 0.251. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.1 are presented in Table 8.13 (page 170). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 99% from 80% (see Figure 8.22). Adjusting for covariate with an ICC of 0.1 increased estimated power to maximum of 100% from 81% (see Figure 8.23, page 169), reducing empirical standard error from 0.096 to 0.03. Adjusting for a cluster level covariate increased estimated power to a maximum of 92% from 81% (see Figure 8.24, page 169), reducing empirical standard error from 0.096 to 0.080.

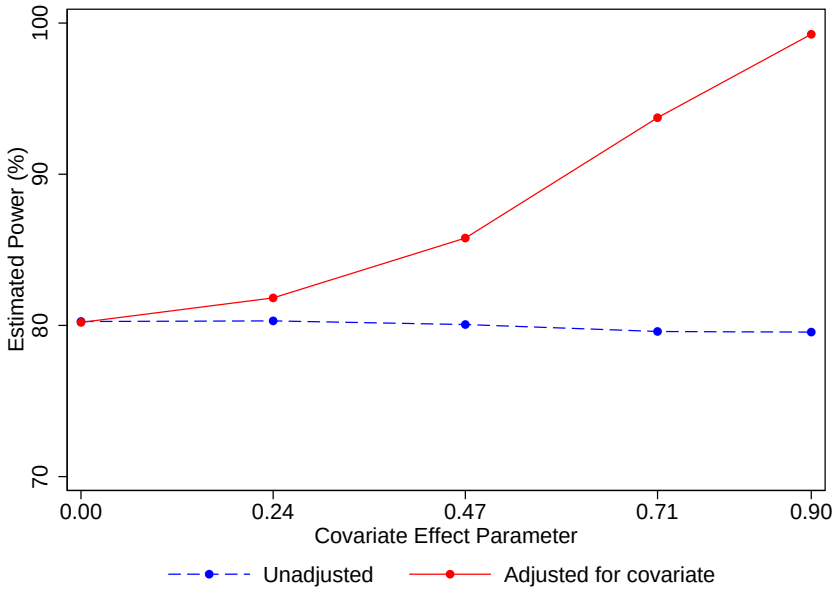


Figure 8.22: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. Covariate ICC = 0.0005. Treatment effect parameter = 0.2706589. Data from Table 8.13.

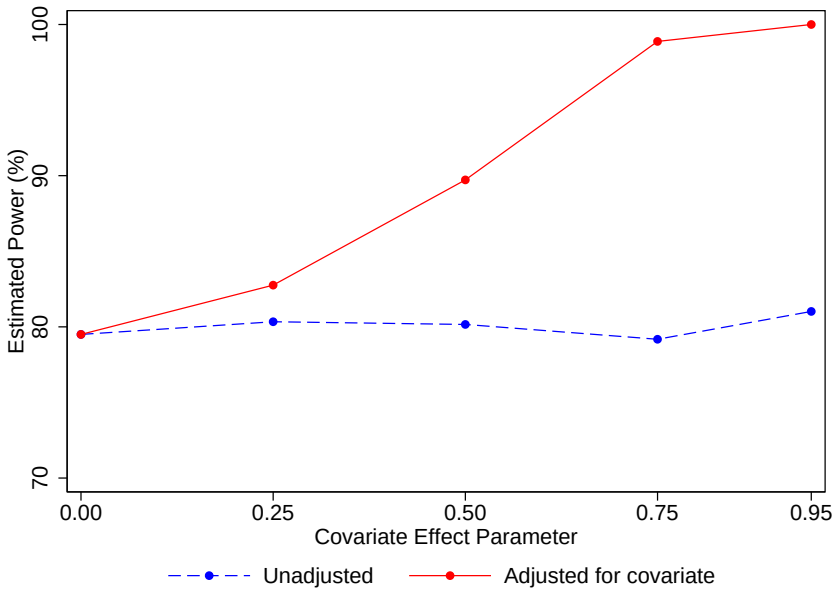


Figure 8.23: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. Covariate ICC = 0.1. Treatment effect parameter = 0.2706589. Data from Table 8.13.

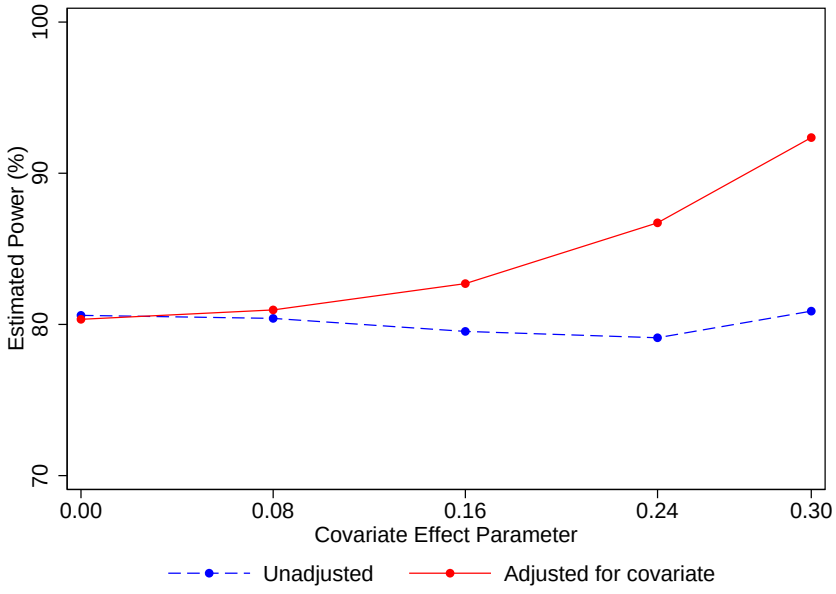


Figure 8.24: Estimated power when using a linear mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. Cluster level covariate. Treatment effect parameter = 0.2706589. Data from Table 8.13.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Power %
Covariate ICC	CEF (parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	From Formula
0.0005	0.00 (0.000)	0.097 (0.001)	0.097 (0.001)	80.3 (0.6)	80.2 (0.6)	79.3
0.0005	0.25 (0.237)	0.096 (0.001)	0.094 (0.001)	80.3 (0.6)	81.8 (0.5)	81.0
0.0005	0.50 (0.474)	0.098 (0.001)	0.090 (0.001)	80.1 (0.6)	85.8 (0.5)	85.9
0.0005	0.75 (0.712)	0.096 (0.001)	0.077 (0.001)	79.6 (0.6)	93.7 (0.3)	93.6
0.0005	0.95 (0.901)	0.096 (0.001)	0.062 (0.001)	79.6 (0.6)	99.3 (0.1)	99.0
0.1	0.00 (0.000)	0.096 (0.001)	0.096 (0.001)	79.5 (0.6)	79.5 (0.6)	79.3
0.1	0.25 (0.250)	0.096 (0.001)	0.094 (0.001)	80.3 (0.6)	82.8 (0.5)	81.9
0.1	0.50 (0.500)	0.096 (0.001)	0.084 (0.001)	80.2 (0.6)	89.7 (0.4)	89.4
0.1	0.75 (0.750)	0.096 (0.001)	0.063 (0.001)	79.2 (0.6)	98.9 (0.1)	98.7
0.1	0.95 (0.950)	0.097 (0.001)	0.030 (<0.001)	81.0 (0.6)	100.0 -	100.0
1	0.00 (0.000)	0.097 (0.001)	0.098 (0.001)	80.6 (0.6)	80.3 (0.6)	79.3
1	0.25 (0.079)	0.096 (0.001)	0.096 (0.001)	80.4 (0.6)	81.0 (0.6)	80.2
1	0.50 (0.158)	0.096 (0.001)	0.093 (0.001)	79.5 (0.6)	82.7 (0.5)	82.9
1	0.75 (0.237)	0.098 (0.001)	0.087 (0.001)	79.1 (0.6)	86.7 (0.5)	87.5
1	0.95 (0.300)	0.096 (0.001)	0.080 (0.001)	80.9 (0.6)	92.4 (0.4)	92.1

Table 8.13: Estimated standard error and power, and power calculated from formula, when using a linear mixed effects model. 60 clusters per treatment arm, cluster size of 5. Outcome ICC = 0.1. Treatment effect parameter = 0.271. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Overall, adjusting for a covariate with the same ICC as the outcome increased estimated power to a maximum of 100%. For an outcome with an ICC of 0.005, adjusting for a covariate with an ICC of 0.0005 also increased estimated power to a maximum of 100%.

With a cluster size of five, adjusting for a covariate with smaller ICC than the outcome generally increased estimated power more than adjusting for a cluster level covariate. With a cluster size of 30, this was true for outcomes with an ICC less than or equal to 0.005. However, for an outcome with an ICC of 0.05 or 0.1, adjusting for a cluster level covariate increased estimated power more than adjusting for a covariate with an ICC of 0.0005. Adjusting for a cluster level covariate was most beneficial for increasing estimated power and reducing empirical standard error when the outcome ICC was larger, and with larger cluster size. The patterns observed when the covariate effect is largest are also seen for smaller covariate effects.

8.4 Effects of covariate adjustment in logistic mixed effects models

In this section I present empirical standard errors and estimated power from simulations of unadjusted and adjusted logistic mixed effects model analyses.

In all situations, adjusting for a covariate with no effect on the outcome did not significantly change the estimated power of the analysis or the empirical standard error of estimated treatment effect. As covariate effect increased, estimated power of adjusted analyses increased.

Adjusting for an individual level covariate typically increased the empirical standard error of the treatment effect estimate. For example, for a CRT with 20 clusters per treatment arm and a cluster size of 30, for an outcome with an ICC of 0.05 and expected value in the control treatment arm of 0.1, adjusting for a covariate with an ICC of 0.05 increased empirical standard error to a maximum of 0.778 from 0.319 (see Table 8.16, page 180, and Figure 8.25, page 172). Adjusting for a cluster level covariate, however, only reduced the empirical standard error. For example, for a CRT with 20 clusters per treatment arm and a cluster size of 30, for an outcome with an ICC of 0.05 and expected value in the control treatment arm of 0.1, adjusting for a cluster level covariate decreased empirical standard error to a minimum of 0.204 from 0.304 (see Table 8.16, page 180, and Figure 8.26, page 172).

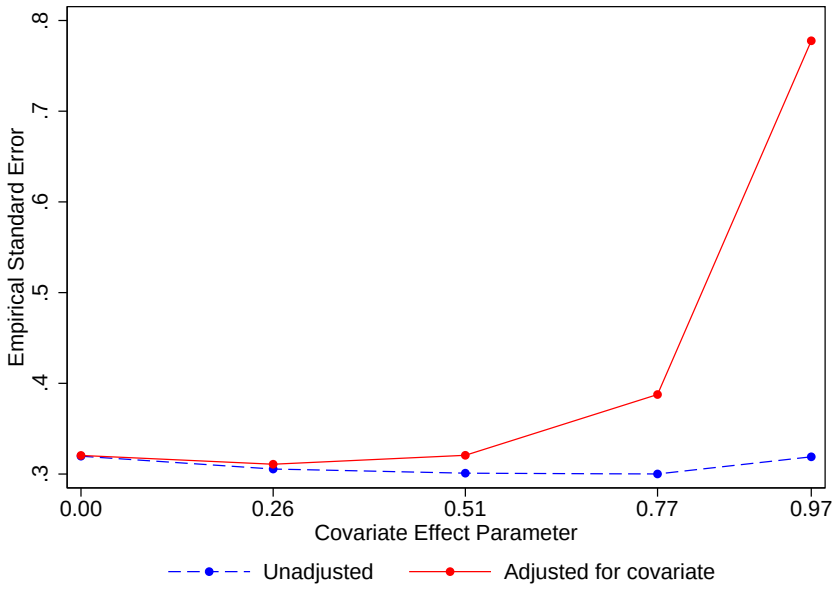


Figure 8.25: Empirical standard error from adjusted logistic mixed effects model analysis. 20 clusters per treatment arm, cluster size of 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Covariate ICC = 0.05. Data from Table 8.16.

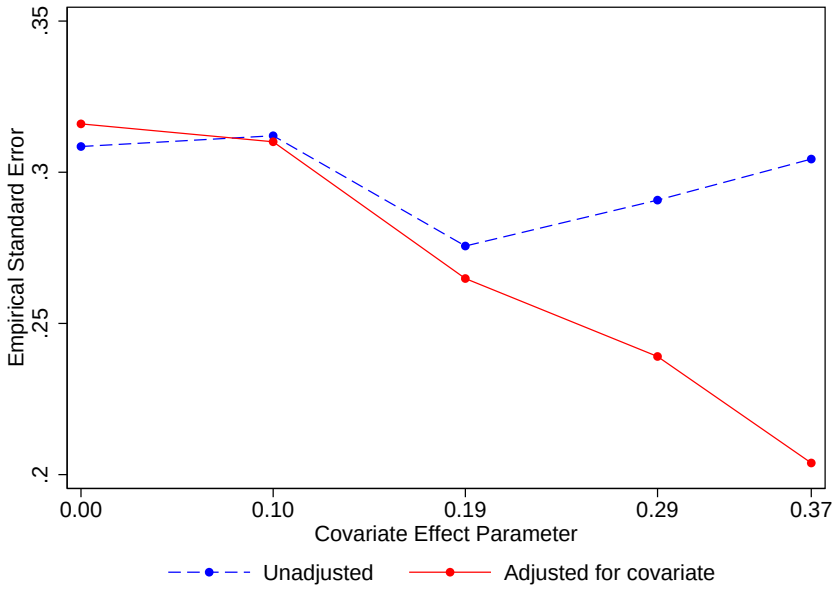


Figure 8.26: Empirical standard error from adjusted logistic mixed effects model analysis. 20 clusters per treatment arm, cluster size of 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Cluster level covariate. Data from Table 8.16.

As discussed in Section 2.2 with regard to individually randomised trials, adjusted and unadjusted treatment effect parameters do not coincide when using logistic regression models. Both are valid measures of treatment effect, but have different interpretations. The unadjusted treatment effect is a marginal effect, while the adjusted treatment effect is a conditional effect. In these results I present the estimated relative asymptotic bias, which is the estimated difference between the adjusted treatment effect estimate and the unadjusted treatment effect parameter.

Estimated relative asymptotic bias was close to zero when adjusting for a cluster level covariate. When adjusting for an individual level covariate, the estimated relative asymptotic bias increased as the covariate effect increased. Although, when adjusting for a covariate that does not notably increase power, the estimates of relative asymptotic bias are small. For example, for a CRT with 60 clusters per treatment arm and a cluster size of five, for an outcome with an ICC of 0.0005 and expected value in the control treatment arm of 0.5, when adjusting for a covariate with an ICC of 0.0005 estimated relative asymptotic bias reached as high as 278% (see Table 8.21, page 195, and Figure 8.27).

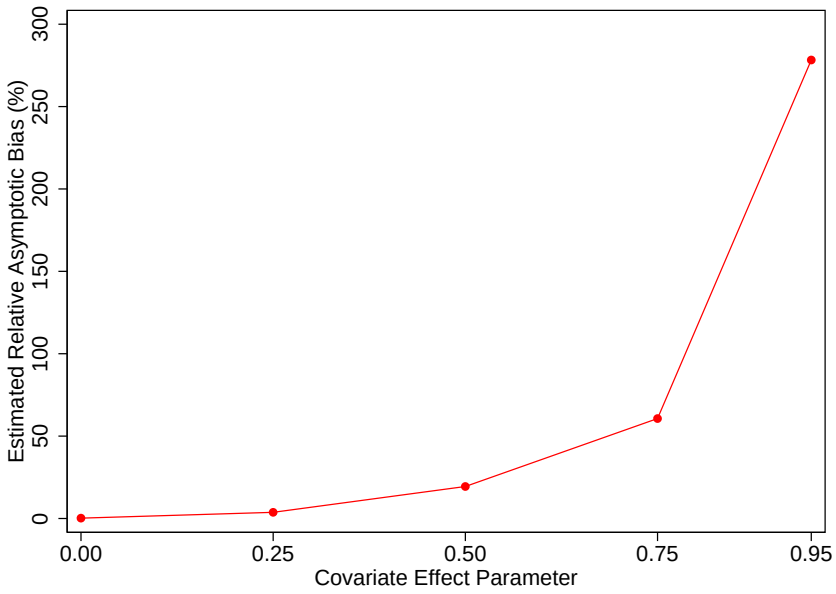


Figure 8.27: Estimated relative asymptotic bias from adjusted logistic mixed effects model analysis. 60 clusters per treatment arm, cluster size of 5. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.0005. Covariate ICC = 0.0005. Data from Table 8.21

Results for simulated CRTs with 20 clusters in each treatment arm and a cluster size of 30 are presented in Tables 8.14 to 8.19.

Results for an outcome with an ICC of 0.0005, and an expected value of the outcome in the control arm of 0.1, are presented in Table 8.14 (page 175). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 100% from 78% (see Figure 8.28). In this case, standard error increased from 0.183 to 0.350 and the estimated relative asymptotic bias was 212.4. Adjusting for a covariate with an ICC of 0.1, or a cluster level covariate, did not significantly change the estimated power or empirical standard error.

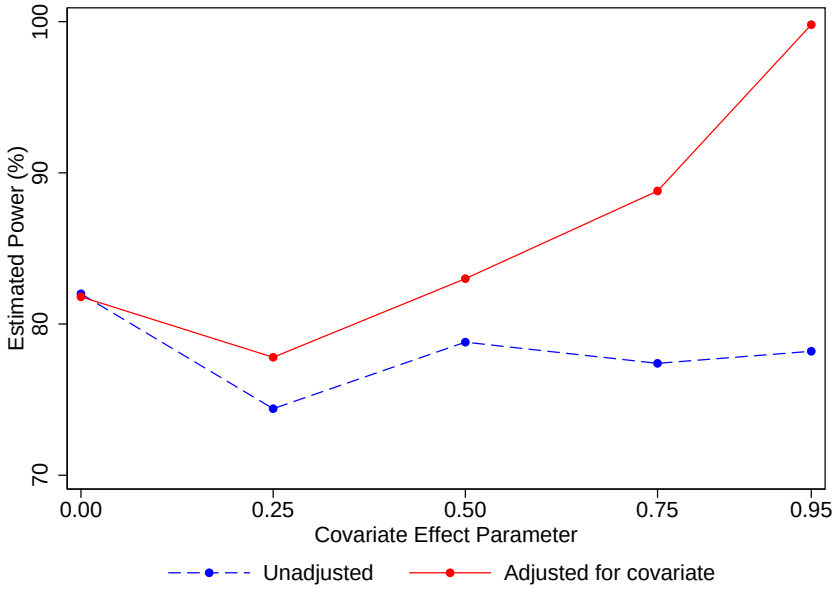


Figure 8.28: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.0005. Covariate ICC = 0.0005. Difference between treatment arms = 0.054. Data from Table 8.14.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.177 (0.006)	0.178 (0.006)	82.0 (1.7)	81.8 (1.7)	0.1
0.0005	0.25 (0.250)	0.189 (0.006)	0.190 (0.006)	74.4 (2.0)	77.8 (1.9)	3.0
0.0005	0.50 (0.500)	0.174 (0.006)	0.189 (0.006)	78.8 (1.8)	83.0 (1.7)	13.5
0.0005	0.75 (0.750)	0.181 (0.006)	0.225 (0.007)	77.4 (1.9)	88.8 (1.4)	49.4
0.0005	0.95 (0.950)	0.183 (0.006)	0.350 (0.011)	78.2 (1.8)	99.8 (0.2)	212.4
0.1	0.00 (0.000)	0.179 (0.006)	0.179 (0.006)	81.6 (1.7)	81.4 (1.7)	0.0
0.1	0.25 (0.030)	0.189 (0.006)	0.189 (0.006)	76.0 (1.9)	76.8 (1.9)	0.1
0.1	0.50 (0.060)	0.175 (0.006)	0.177 (0.006)	80.2 (1.8)	80.4 (1.8)	0.0
0.1	0.75 (0.091)	0.182 (0.006)	0.181 (0.006)	78.2 (1.8)	79.6 (1.8)	0.5
0.1	0.95 (0.115)	0.184 (0.006)	0.183 (0.006)	76.6 (1.9)	77.8 (1.9)	0.3
1	0.00 (0.000)	0.178 (0.006)	0.181 (0.006)	77.4 (1.9)	77.0 (1.9)	0.3
1	0.25 (0.010)	0.176 (0.006)	0.180 (0.006)	77.8 (1.9)	76.6 (1.9)	0.3
1	0.50 (0.019)	0.180 (0.006)	0.184 (0.006)	77.6 (1.9)	76.6 (1.9)	0.0
1	0.75 (0.029)	0.175 (0.006)	0.178 (0.006)	81.0 (1.8)	80.6 (1.8)	−0.1
1	0.95 (0.036)	0.176 (0.006)	0.179 (0.006)	78.2 (1.8)	78.4 (1.8)	−0.1

Table 8.14: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 10 clusters per treatment arm, cluster size 30. Outcome ICC =0.0005. Difference between treatment arms = 0.054. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.0005, and an expected value of the outcome in the control arm of 0.5, are presented in Table 8.15 (page 177). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 100% from 79% (see Figure 8.29). Empirical standard error increased from 0.111 to 0.239, and the estimated relative asymptotic bias was 259%. Adjusting for a covariate with an ICC of 0.1, or a cluster level covariate, did not significantly change the estimated power or empirical standard error.

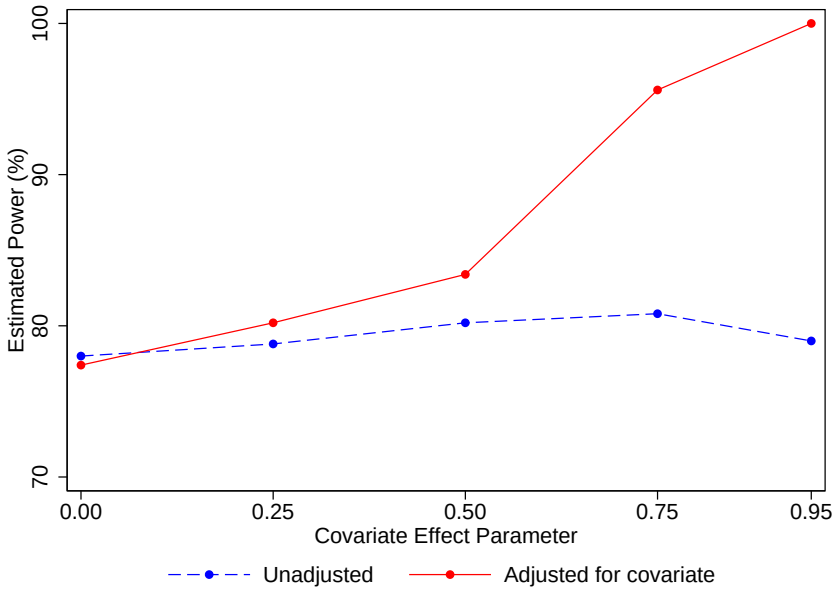


Figure 8.29: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.0005. Covariate ICC = 0.0005. Difference between treatment arms = 0.081. Data from Table 8.15.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.116 (0.004)	0.117 (0.004)	78.0 (1.9)	77.4 (1.9)	0.1
0.0005	0.25 (0.250)	0.115 (0.004)	0.118 (0.004)	78.8 (1.8)	80.2 (1.8)	4.4
0.0005	0.50 (0.500)	0.115 (0.004)	0.127 (0.004)	80.2 (1.8)	83.4 (1.7)	18.3
0.0005	0.75 (0.750)	0.115 (0.004)	0.150 (0.005)	80.8 (1.8)	95.6 (0.9)	63.8
0.0005	0.95 (0.950)	0.111 (0.004)	0.239 (0.008)	79.0 (1.8)	100.0 -	259.4
0.1	0.00 (0.000)	0.118 (0.004)	0.117 (0.004)	75.2 (1.9)	74.8 (1.9)	0.2
0.1	0.25 (0.022)	0.114 (0.004)	0.114 (0.004)	79.0 (1.8)	78.6 (1.8)	0.1
0.1	0.50 (0.044)	0.121 (0.004)	0.121 (0.004)	76.6 (1.9)	75.8 (1.9)	0.2
0.1	0.75 (0.066)	0.118 (0.004)	0.119 (0.004)	77.0 (1.9)	77.0 (1.9)	0.7
0.1	0.95 (0.084)	0.121 (0.004)	0.121 (0.004)	78.4 (1.8)	78.2 (1.8)	0.6
1	0.00 (0.000)	0.122 (0.004)	0.124 (0.004)	75.6 (1.9)	74.4 (2.0)	−0.1
1	0.25 (0.007)	0.119 (0.004)	0.120 (0.004)	76.2 (1.9)	76.6 (1.9)	0.3
1	0.50 (0.014)	0.114 (0.004)	0.117 (0.004)	80.8 (1.8)	79.0 (1.8)	0.2
1	0.75 (0.021)	0.118 (0.004)	0.119 (0.004)	76.0 (1.9)	75.0 (1.9)	−0.1
1	0.95 (0.027)	0.119 (0.004)	0.120 (0.004)	78.2 (1.8)	77.6 (1.9)	−0.1

Table 8.15: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.0005. Difference between treatment arms = 0.081. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.05, and an expected value of 0.1 in the control treatment arm, are presented in Table 8.16 (page 180). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 87% from 79% (see Figure 8.30). Empirical standard error increased from 0.280 to 0.794, and the estimated relative asymptotic bias was 219%. Adjusting for a covariate with an ICC of 0.05 increased estimated power to a maximum of 94% from 77% achieved with the unadjusted analysis (see Figure 8.31, 179). In this case, empirical standard error increased from 0.319 to 0.778, and the estimated relative asymptotic bias was 214%. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 99.8% from 81%. Empirical standard error increased from 0.288 to 0.598, and the estimated relative asymptotic bias was 216%. Adjusting for a cluster level covariate increased estimated power to a maximum of 99% from 77% (see Figure 8.32, page 179). Here, empirical standard error reduced from 0.304 to 0.204.

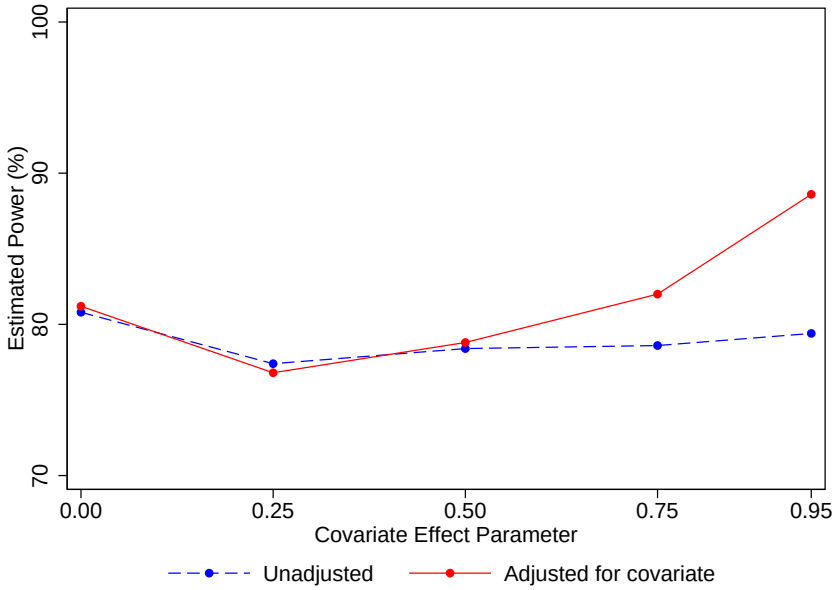


Figure 8.30: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Covariate ICC = 0.0005. Difference between treatment arms = 0.088. Data from Table 8.16.

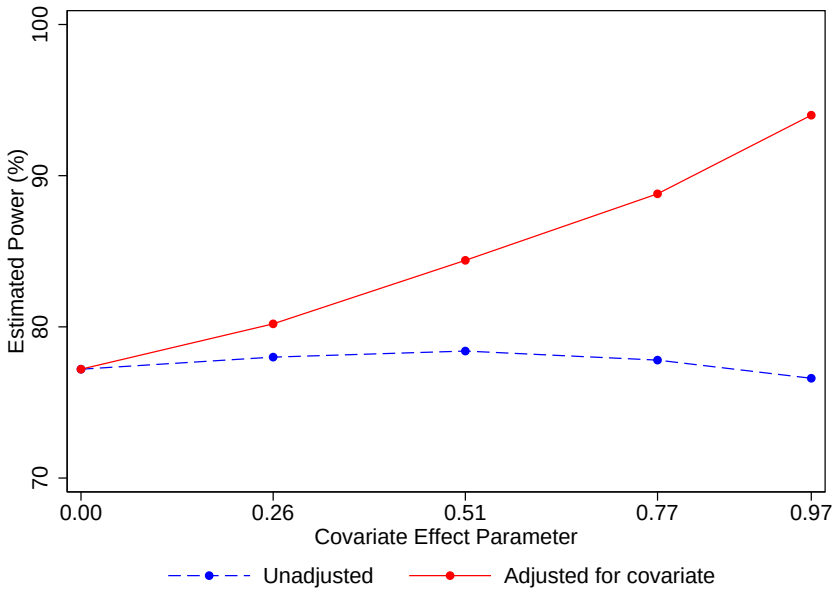


Figure 8.31: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Covariate ICC = 0.05. Difference between treatment arms = 0.088. Data from Table 8.16.

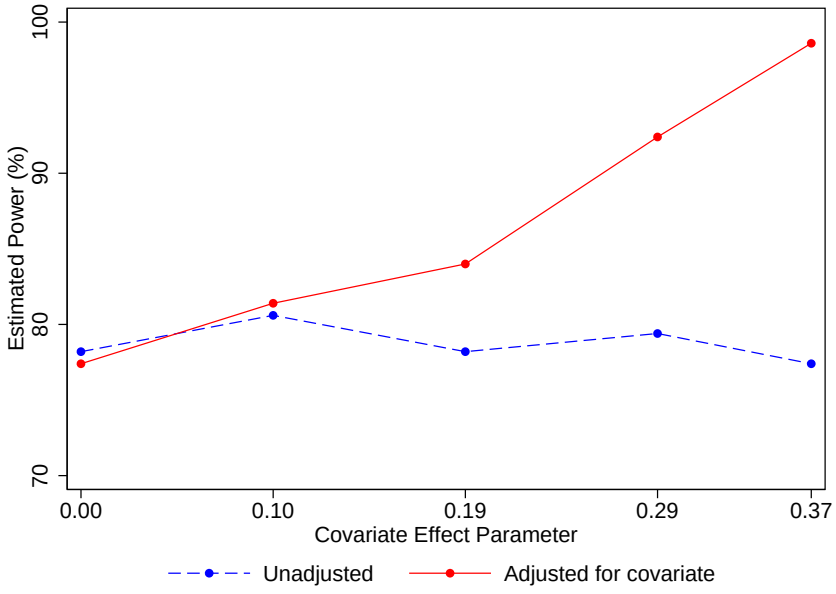


Figure 8.32: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Cluster level covariate. Difference between treatment arms = 0.088. Data from Table 8.16.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.289 (0.009)	0.290 (0.009)	80.8 (1.8)	81.2 (1.7)	0.1
0.0005	0.25 (0.250)	0.290 (0.009)	0.297 (0.009)	77.4 (1.9)	76.8 (1.9)	3.2
0.0005	0.50 (0.500)	0.291 (0.009)	0.328 (0.010)	78.4 (1.8)	78.8 (1.8)	14.1
0.0005	0.75 (0.750)	0.311 (0.010)	0.439 (0.014)	78.6 (1.8)	82.0 (1.7)	47.7
0.0005	0.95 (0.950)	0.280 (0.009)	0.794 (0.025)	79.4 (1.8)	88.6 (1.4)	219.3
0.05	0.00 (0.000)	0.320 (0.010)	0.321 (0.010)	77.2 (1.9)	77.2 (1.9)	0.2
0.05	0.25 (0.256)	0.306 (0.010)	0.311 (0.010)	78.0 (1.9)	80.2 (1.8)	2.9
0.05	0.50 (0.513)	0.301 (0.010)	0.321 (0.010)	78.4 (1.8)	84.4 (1.6)	14.1
0.05	0.75 (0.769)	0.300 (0.009)	0.388 (0.012)	77.8 (1.9)	88.8 (1.4)	49.2
0.05	0.95 (0.975)	0.319 (0.010)	0.778 (0.025)	76.6 (1.9)	94.0 (1.1)	214.0

Continued on next page.

Table 8.16: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Difference between treatment arms = 0.088. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.1	0.00 (0.000)	0.295 (0.009)	0.296 (0.009)	79.6 (1.8)	79.8 (1.8)	0.2
0.1	0.25 (0.264)	0.306 (0.010)	0.310 (0.010)	80.2 (1.8)	81.6 (1.7)	2.7
0.1	0.50 (0.527)	0.297 (0.009)	0.306 (0.010)	77.4 (1.9)	84.4 (1.6)	13.9
0.1	0.75 (0.791)	0.289 (0.009)	0.356 (0.011)	78.6 (1.8)	92.8 (1.2)	47.7
0.1	0.95 (1.001)	0.288 (0.009)	0.598 (0.019)	81.0 (1.8)	99.8 (0.2)	216.1
1	0.00 (0.000)	0.309 (0.010)	0.316 (0.010)	78.2 (1.8)	77.4 (1.9)	0.0
1	0.25 (0.097)	0.312 (0.010)	0.310 (0.010)	80.6 (1.8)	81.4 (1.7)	0.6
1	0.50 (0.195)	0.276 (0.009)	0.265 (0.008)	78.2 (1.8)	84.0 (1.6)	0.8
1	0.75 (0.292)	0.291 (0.009)	0.239 (0.008)	79.4 (1.8)	92.4 (1.2)	0.7
1	0.95 (0.370)	0.304 (0.010)	0.204 (0.006)	77.4 (1.9)	98.6 (0.5)	1.1

Table 8.16: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Difference between treatment arms = 0.088. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.05, and an expected value of 0.5 in the control treatment arm, are presented in Table 8.17 (page 184). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 91% from 83% (see Figure 8.33), while empirical standard error increased from 0.194 to 0.602. Adjusting for a covariate with an ICC of 0.05 increased estimated power to a maximum of 99% from 82% (see Figure 8.34, page 183). Empirical standard error increased from 0.191 to 0.461, and estimated relative asymptotic bias was 252%. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 100% from 77%. Adjusting for a cluster level covariate increased estimated power to a maximum of 98% from 82% (see Figure 8.35, page 183). Here, empirical standard error reduced from 0.188 to 0.132.

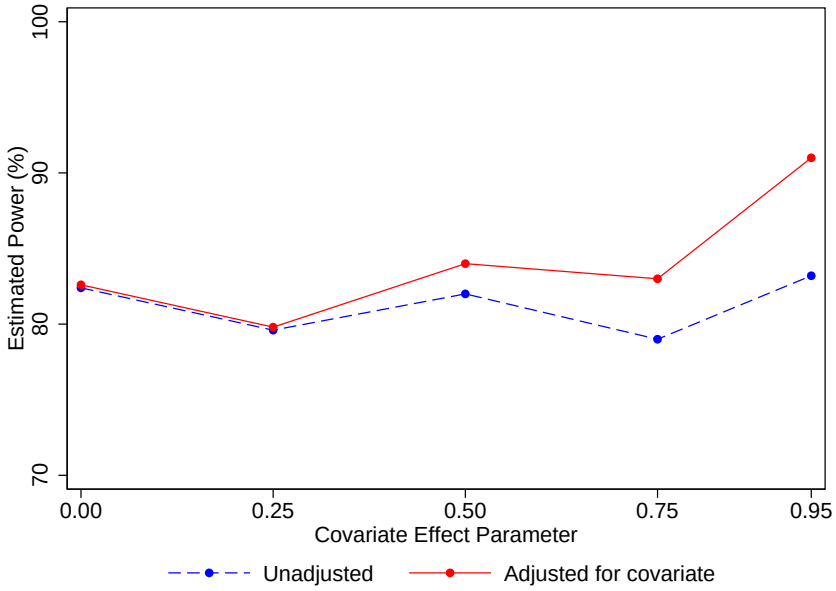


Figure 8.33: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.05. Covariate ICC = 0.0005. Difference between treatment arms = 0.125. Data from Table 8.17.

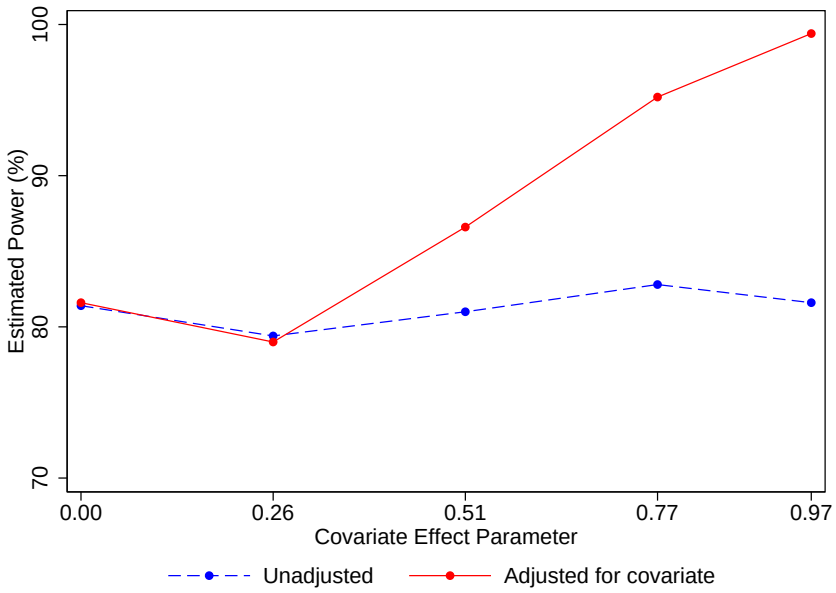


Figure 8.34: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.05. Covariate ICC = 0.05. Difference between treatment arms = 0.125. Data from Table 8.17.

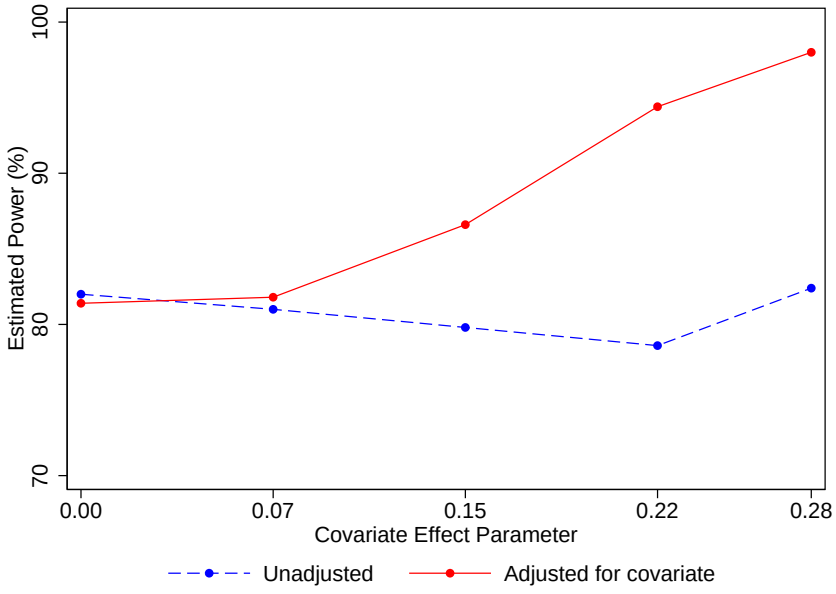


Figure 8.35: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.05. Cluster level covariate. Difference between treatment arms = 0.125. Data from Table 8.17.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.199 (0.006)	0.199 (0.006)	82.4 (1.7)	82.6 (1.7)	0.1
0.0005	0.25 (0.250)	0.192 (0.006)	0.198 (0.006)	79.6 (1.8)	79.8 (1.8)	3.8
0.0005	0.50 (0.500)	0.194 (0.006)	0.224 (0.007)	82.0 (1.7)	84.0 (1.6)	18.7
0.0005	0.75 (0.750)	0.200 (0.006)	0.295 (0.009)	79.0 (1.8)	83.0 (1.7)	61.0
0.0005	0.95 (0.950)	0.194 (0.006)	0.602 (0.019)	83.2 (1.7)	91.0 (1.3)	257.4
0.05	0.00 (0.000)	0.194 (0.006)	0.194 (0.006)	81.4 (1.7)	81.6 (1.7)	0.1
0.05	0.25 (0.256)	0.192 (0.006)	0.195 (0.006)	79.4 (1.8)	79.0 (1.8)	4.0
0.05	0.50 (0.513)	0.198 (0.006)	0.212 (0.007)	81.0 (1.8)	86.6 (1.5)	20.0
0.05	0.75 (0.769)	0.181 (0.006)	0.234 (0.007)	82.8 (1.7)	95.2 (1.0)	61.7
0.05	0.95 (0.975)	0.191 (0.006)	0.461 (0.015)	81.6 (1.7)	99.4 (0.3)	251.6

Continued on next page.

Table 8.17: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Difference between treatment arms = 0.125. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.1	0.00 (0.000)	0.206 (0.007)	0.206 (0.007)	79.6 (1.8)	79.2 (1.8)	0.1
0.1	0.25 (0.231)	0.197 (0.006)	0.197 (0.006)	79.8 (1.8)	82.0 (1.7)	3.2
0.1	0.50 (0.461)	0.195 (0.006)	0.199 (0.006)	78.8 (1.8)	88.2 (1.4)	13.9
0.1	0.75 (0.692)	0.198 (0.006)	0.204 (0.006)	79.8 (1.8)	96.6 (0.8)	38.9
0.1	0.95 (0.877)	0.196 (0.006)	0.196 (0.006)	76.8 (1.9)	100.0 -	99.2
1	0.00 (0.000)	0.200 (0.006)	0.202 (0.006)	82.0 (1.7)	81.4 (1.7)	−0.2
1	0.25 (0.073)	0.199 (0.006)	0.194 (0.006)	81.0 (1.8)	81.8 (1.7)	−0.2
1	0.50 (0.146)	0.186 (0.006)	0.176 (0.006)	79.8 (1.8)	86.6 (1.5)	0.2
1	0.75 (0.219)	0.192 (0.006)	0.157 (0.005)	78.6 (1.8)	94.4 (1.0)	−0.4
1	0.95 (0.277)	0.188 (0.006)	0.132 (0.004)	82.4 (1.7)	98.0 (0.6)	0.3

Table 8.17: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Difference between treatment arms = 0.125. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.1, and an expected value of 0.1 in the control treatment arm, are presented in Table 8.18 (page 188). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 89% from 81% (see Figure 8.36), while empirical standard error increased from 0.386 to 1.136. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 93% from 81% (see Figure 8.37, page 187). Empirical standard error increased from 0.381 to 1.024, and estimated relative asymptotic bias was 215%. Adjusting for a cluster level covariate increased estimated power to a maximum of 100% from 81% (see Figure 8.38, page 187). Here, empirical standard error reduced from 0.188 to 0.132.

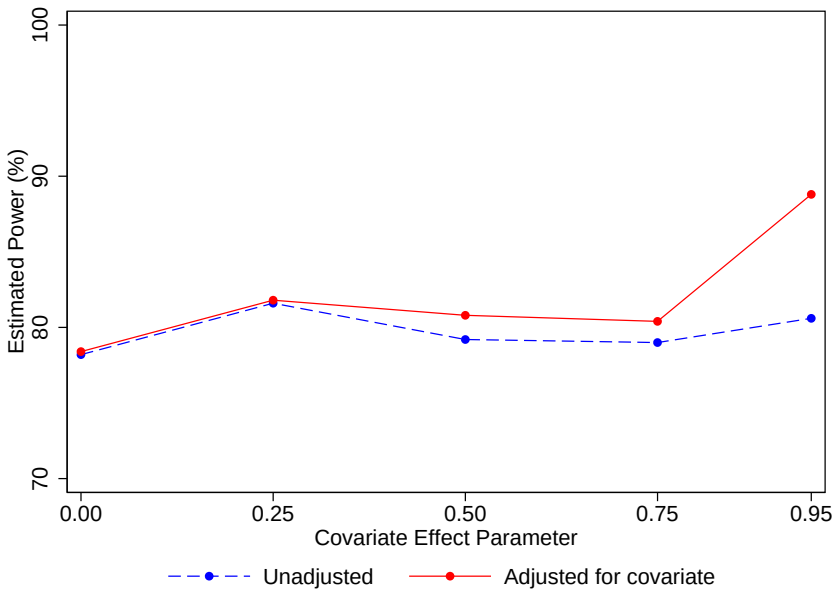


Figure 8.36: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.1. Covariate ICC = 0.0005. Difference between treatment arms = 0.115. Data from Table 8.18.

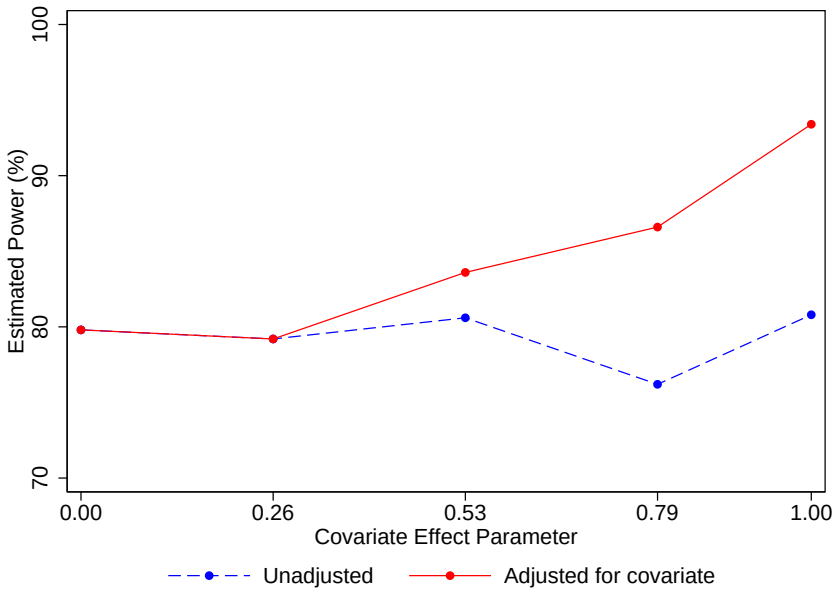


Figure 8.37: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.1. Covariate ICC = 0.1. Difference between treatment arms = 0.115. Data from Table 8.18.

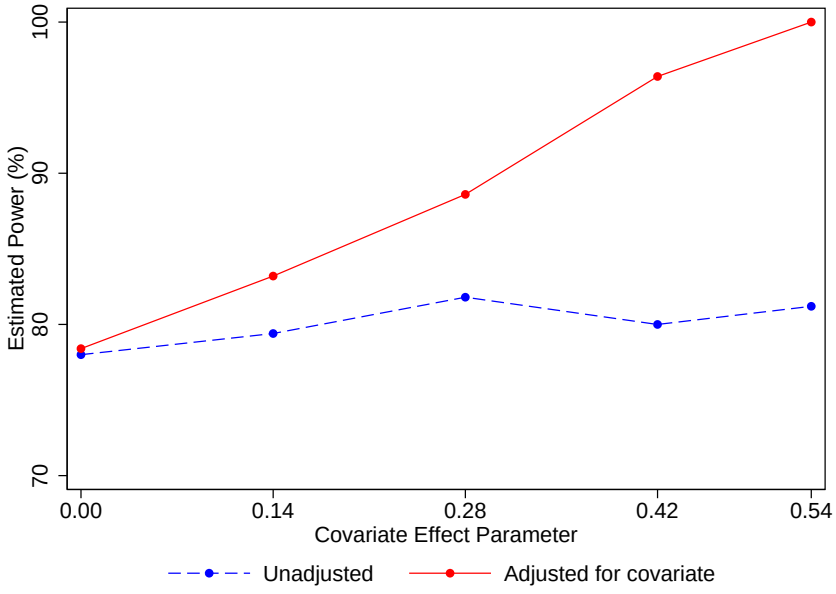


Figure 8.38: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.1. Cluster level covariate. Difference between treatment arms = 0.115. Data from Table 8.18.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.396 (0.013)	0.397 (0.013)	78.2 (1.8)	78.4 (1.8)	0.1
0.0005	0.25 (0.250)	0.386 (0.012)	0.396 (0.013)	81.6 (1.7)	81.8 (1.7)	3.1
0.0005	0.50 (0.500)	0.398 (0.013)	0.444 (0.014)	79.2 (1.8)	80.8 (1.8)	14.5
0.0005	0.75 (0.750)	0.395 (0.013)	0.573 (0.018)	79.0 (1.8)	80.4 (1.8)	50.2
0.0005	0.95 (0.950)	0.386 (0.012)	1.136 (0.036)	80.6 (1.8)	88.8 (1.4)	218.2
0.1	0.00 (0.000)	0.379 (0.012)	0.379 (0.012)	79.8 (1.8)	79.8 (1.8)	0.1
0.1	0.25 (0.264)	0.391 (0.012)	0.401 (0.013)	79.2 (1.8)	79.2 (1.8)	2.8
0.1	0.50 (0.527)	0.382 (0.012)	0.417 (0.013)	80.6 (1.8)	83.6 (1.7)	14.2
0.1	0.75 (0.791)	0.382 (0.012)	0.502 (0.016)	76.2 (1.9)	86.6 (1.5)	48.9
0.1	0.95 (1.001)	0.381 (0.012)	1.024 (0.032)	80.8 (1.8)	93.4 (1.1)	214.6
1	0.00 (0.000)	0.395 (0.012)	0.402 (0.013)	78.0 (1.9)	78.4 (1.8)	−0.1
1	0.25 (0.141)	0.384 (0.012)	0.375 (0.012)	79.4 (1.8)	83.2 (1.7)	0.2
1	0.50 (0.283)	0.376 (0.012)	0.340 (0.011)	81.8 (1.7)	88.6 (1.4)	−0.6
1	0.75 (0.424)	0.385 (0.012)	0.298 (0.009)	80.0 (1.8)	96.4 (0.8)	−1.4
1	0.95 (0.537)	0.377 (0.012)	0.221 (0.007)	81.2 (1.7)	100.0 -	0.3

Table 8.18: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. Difference between treatment arms = 0.115. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.1, and an expected value of 0.5 in the control treatment arm, are presented in Table 8.19 (page 191). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 89% from 83% (see Figure 8.39), while empirical standard error increased from 0.253 to 0.828. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 99% from 79% (see Figure 8.40, page 190). Empirical standard error increased from 0.261 to 0.602, and estimated relative asymptotic bias was 258%. Adjusting for a cluster level covariate increased estimated power to a maximum of 100% from 83% (see Figure 8.41, page 190). Here, empirical standard error reduced from 0.271 to 0.146.

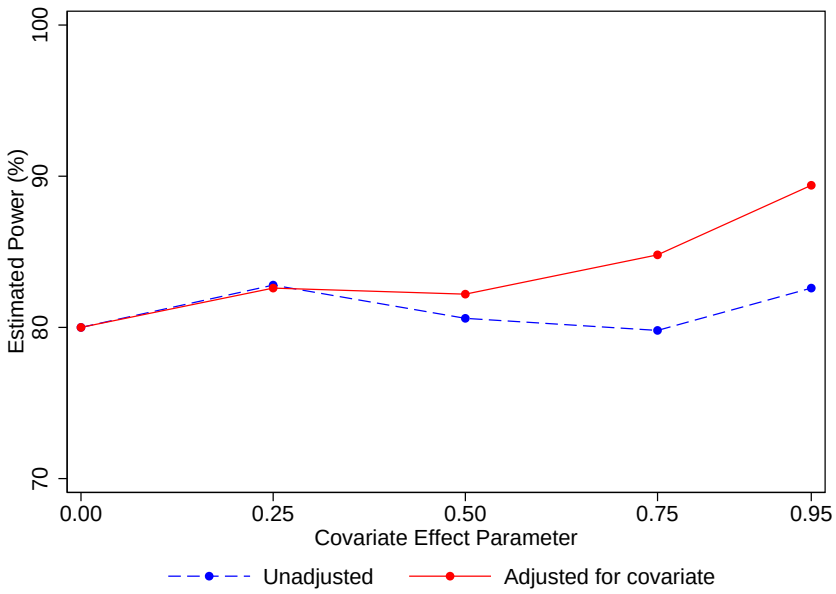


Figure 8.39: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.1. Covariate ICC = 0.0005. Difference between treatment arms = 0.156. Data from Table 8.19.

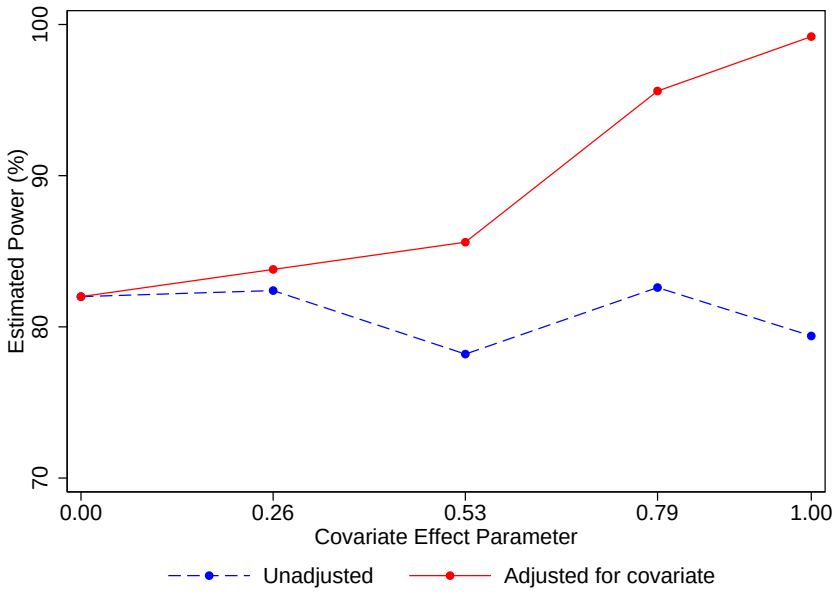


Figure 8.40: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.1. Covariate ICC = 0.1. Difference between treatment arms = 0.156. Data from Table 8.19.

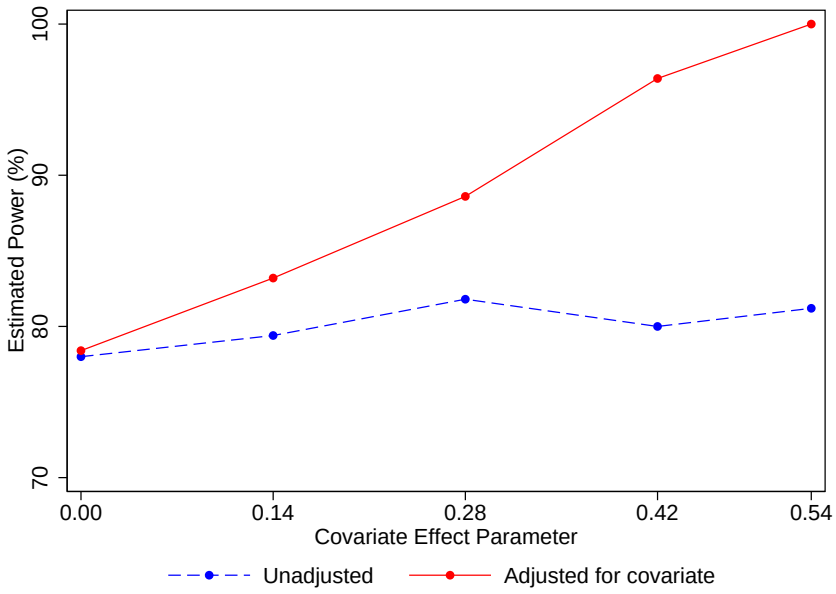


Figure 8.41: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 20 clusters per treatment arm, cluster size 30. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.1. Cluster level covariate. Difference between treatment arms = 0.156. Data from Table 8.19.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.248 (0.008)	0.248 (0.008)	80.0 (1.8)	80.0 (1.8)	0.1
0.0005	0.25 (0.250)	0.262 (0.008)	0.271 (0.009)	82.8 (1.7)	82.6 (1.7)	3.9
0.0005	0.50 (0.500)	0.255 (0.008)	0.299 (0.009)	80.6 (1.8)	82.2 (1.7)	18.6
0.0005	0.75 (0.750)	0.268 (0.008)	0.393 (0.012)	79.8 (1.8)	84.8 (1.6)	59.6
0.0005	0.95 (0.950)	0.253 (0.008)	0.828 (0.026)	82.6 (1.7)	89.4 (1.4)	253.7
0.1	0.00 (0.000)	0.248 (0.008)	0.248 (0.008)	82.0 (1.7)	82.0 (1.7)	0.1
0.1	0.25 (0.264)	0.263 (0.008)	0.271 (0.009)	82.4 (1.7)	83.8 (1.6)	4.0
0.1	0.50 (0.527)	0.247 (0.008)	0.271 (0.009)	78.2 (1.8)	85.6 (1.6)	17.7
0.1	0.75 (0.791)	0.260 (0.008)	0.316 (0.010)	82.6 (1.7)	95.6 (0.9)	59.4
0.1	0.95 (1.001)	0.261 (0.008)	0.602 (0.019)	79.4 (1.8)	99.2 (0.4)	257.9
1	0.00 (0.000)	0.253 (0.008)	0.257 (0.008)	83.6 (1.7)	84.0 (1.6)	0.1
1	0.25 (0.108)	0.251 (0.008)	0.255 (0.008)	81.0 (1.8)	80.2 (1.8)	0.0
1	0.50 (0.215)	0.249 (0.008)	0.229 (0.007)	82.0 (1.7)	89.4 (1.4)	0.0
1	0.75 (0.323)	0.264 (0.008)	0.201 (0.006)	79.6 (1.8)	95.6 (0.9)	−1.2
1	0.95 (0.409)	0.271 (0.009)	0.146 (0.005)	82.8 (1.7)	100.0 -	−0.6

Table 8.19: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. Difference between treatment arms = 0.156. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Empirical standard error and estimated power under unadjusted and adjusted analyses are presented in Tables 8.20 to 8.25 (pages 193 to 209) for simulated CRTs with 60 clusters in each treatment arm and a cluster size of five. The tables also present the estimated relative asymptotic bias.

Results for an outcome with an ICC of 0.0005, and an expected value of the outcome in the control arm of 0.1, are presented in Table 8.20 (page 193). Adjusting for a covariate with ICC of 0.0005 increased estimated power to a maximum of 99% from 77% achieved with the unadjusted model (see Figure 8.42). In this case, empirical standard error increased from 0.268 to 0.718, and the estimate relative asymptotic bias was 241%. Adjusting for a covariate with ICC of 0.05 or 1 did not significantly increase estimated power from the unadjusted analysis.

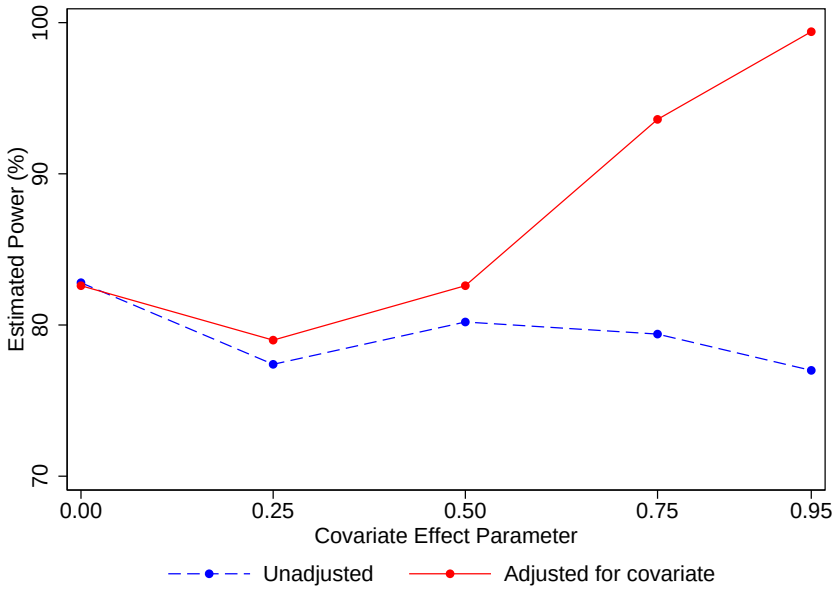


Figure 8.42: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.0005. Covariate ICC = 0.0005. Difference between treatment arms = 0.079. Data from Table 8.20.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.230 (0.007)	0.230 (0.007)	82.8 (1.7)	82.6 (1.7)	0.2
0.0005	0.25 (0.250)	0.265 (0.008)	0.267 (0.008)	77.4 (1.9)	79.0 (1.8)	3.1
0.0005	0.50 (0.500)	0.258 (0.008)	0.283 (0.009)	80.2 (1.8)	82.6 (1.7)	14.8
0.0005	0.75 (0.750)	0.254 (0.008)	0.318 (0.010)	79.4 (1.8)	93.6 (1.1)	50.7
0.0005	0.95 (0.950)	0.268 (0.008)	0.718 (0.023)	77.0 (1.9)	99.4 (0.3)	241.4
0.1	0.00 (0.000)	0.251 (0.008)	0.252 (0.008)	81.4 (1.7)	80.8 (1.8)	0.2
0.1	0.25 (0.030)	0.262 (0.008)	0.263 (0.008)	76.6 (1.9)	77.0 (1.9)	0.3
0.1	0.50 (0.060)	0.255 (0.008)	0.255 (0.008)	79.6 (1.8)	79.8 (1.8)	0.4
0.1	0.75 (0.091)	0.259 (0.008)	0.260 (0.008)	79.8 (1.8)	80.6 (1.8)	0.6
0.1	0.95 (0.115)	0.258 (0.008)	0.261 (0.008)	76.0 (1.9)	75.8 (1.9)	0.8
1	0.00 (0.000)	0.249 (0.008)	0.251 (0.008)	78.6 (1.8)	78.8 (1.8)	0.2
1	0.25 (0.010)	0.237 (0.007)	0.237 (0.008)	81.2 (1.7)	80.6 (1.8)	−0.2
1	0.50 (0.019)	0.263 (0.008)	0.265 (0.008)	75.8 (1.9)	75.2 (1.9)	0.1
1	0.75 (0.029)	0.258 (0.008)	0.257 (0.008)	79.2 (1.8)	79.6 (1.8)	0.4
1	0.95 (0.036)	0.261 (0.008)	0.263 (0.008)	76.8 (1.91)	75.8 (1.93)	0.1

Table 8.20: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.0005. Difference between treatment arms = 0.079. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.0005, and an expected value of the outcome in the control arm of 0.5, are presented in Table 8.21 (page 195). Adjusting for a covariate with ICC of 0.0005 increased estimated power to a maximum of 100% from 81% achieved with the unadjusted model (see Figure 8.43). In this case, empirical standard error increased from 0.168 to 0.380, and the estimated relative asymptotic bias was 278%. Adjusting for a covariate with ICC of 0.05 or 1 did not significantly increase estimated power from the unadjusted analysis.

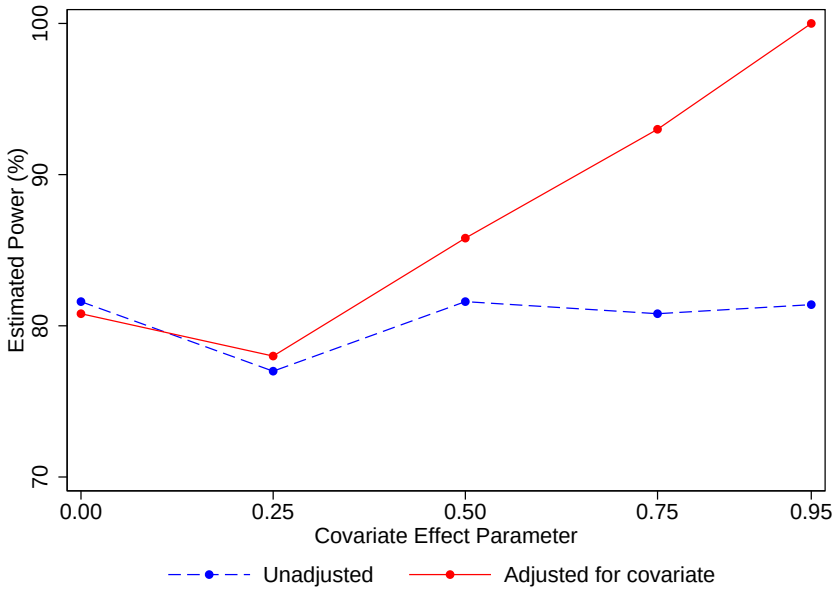


Figure 8.43: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.0005. Covariate ICC = 0.0005. Difference between treatment arms = 0.079. Data from Table 8.21.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.167 (0.005)	0.168 (0.005)	81.6 (1.7)	80.8 (1.8)	0.2
0.0005	0.25 (0.250)	0.176 (0.006)	0.181 (0.006)	77.0 (1.9)	78.0 (1.9)	3.8
0.0005	0.50 (0.500)	0.159 (0.005)	0.175 (0.006)	81.6 (1.7)	85.8 (1.6)	19.4
0.0005	0.75 (0.750)	0.167 (0.005)	0.220 (0.007)	80.8 (1.8)	93.0 (1.1)	60.7
0.0005	0.95 (0.950)	0.168 (0.005)	0.380 (0.012)	81.4 (1.7)	100.0 (0.00)	278.3
0.1	0.00 (0.000)	0.172 (0.005)	0.172 (0.005)	78.6 (1.8)	78.6 (1.8)	0.2
0.1	0.25 (0.022)	0.165 (0.005)	0.165 (0.005)	77.2 (1.9)	77.4 (1.9)	0.0
0.1	0.50 (0.044)	0.171 (0.005)	0.172 (0.005)	75.8 (1.9)	75.2 (1.9)	0.3
0.1	0.75 (0.066)	0.170 (0.005)	0.170 (0.005)	81.4 (1.7)	81.4 (1.7)	0.4
0.1	0.95 (0.084)	0.161 (0.005)	0.162 (0.005)	78.6 (1.8)	78.8 (1.8)	0.3
1	0.00 (0.000)	0.164 (0.005)	0.164 (0.005)	80.8 (1.8)	80.4 (1.8)	0.1
1	0.25 (0.007)	0.172 (0.005)	0.174 (0.006)	79.4 (1.8)	78.8 (1.8)	0.6
1	0.50 (0.014)	0.173 (0.005)	0.175 (0.006)	77.6 (1.9)	77.2 (1.9)	0.0
1	0.75 (0.021)	0.170 (0.005)	0.170 (0.005)	77.0 (1.9)	76.2 (1.9)	−0.1
1	0.95 (0.027)	0.179 (0.006)	0.178 (0.006)	78.6 (1.8)	78.6 (1.8)	0.2

Table 8.21: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.0005. Difference between treatment arms = 0.113. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.05, and an expected value of the outcome in the control arm of 0.1, are presented in Table 8.22 (page 198). Adjusting for a covariate with ICC of 0.0005 increased power to a maximum of 93% from 78% achieved with the unadjusted model (see Figure 8.44). Empirical standard error increase from 0.287 to 0.884, and the estimated relative asymptotic bias was 235%. Adjusting for a covariate with ICC of 0.05 increased estimated power to a maximum of 97% from 78% achieved with the unadjusted model (see Figure 8.45, page 197). In this case, empirical standard error increased from 0.309 to 0.855, and the estimated relative asymptotic bias was 233%. Adjusting for a cluster level covariate increased estimated power to a maximum of 85% from 76% achieved with the unadjusted model (see Figure 8.46, 197), and empirical standard error decreased from 0.293 to 0.267.

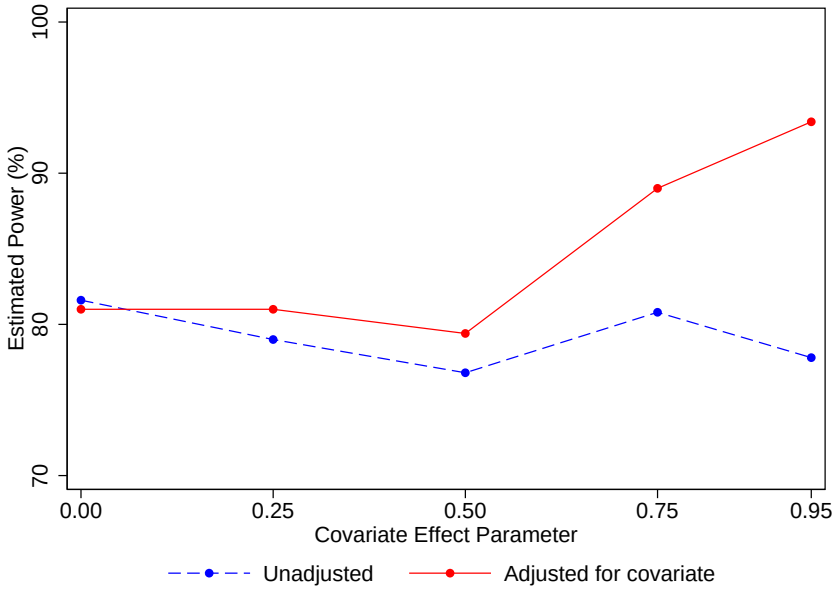


Figure 8.44: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Covariate ICC = 0.0005. Difference between treatment arms = 0.087. Data from Table 8.22.

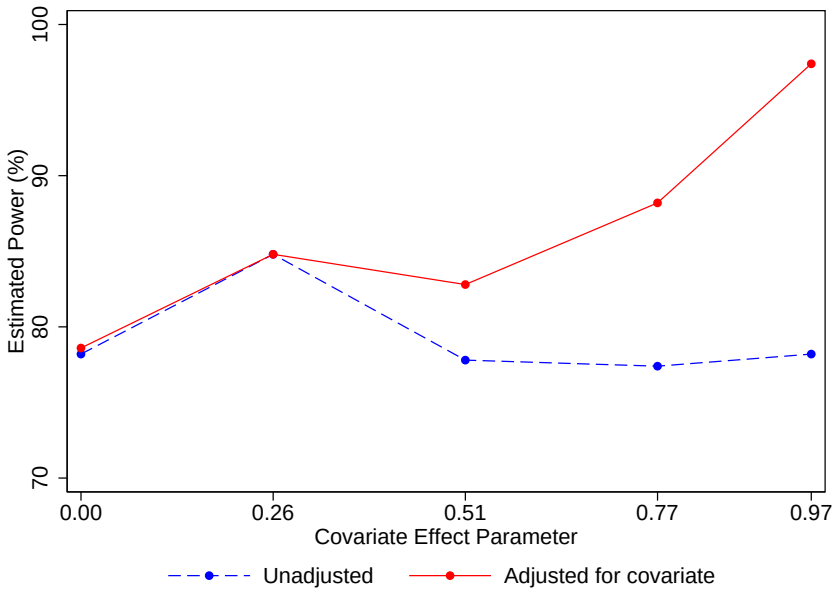


Figure 8.45: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Covariate ICC = 0.05. Difference between treatment arms = 0.087. Data from Table 8.22.

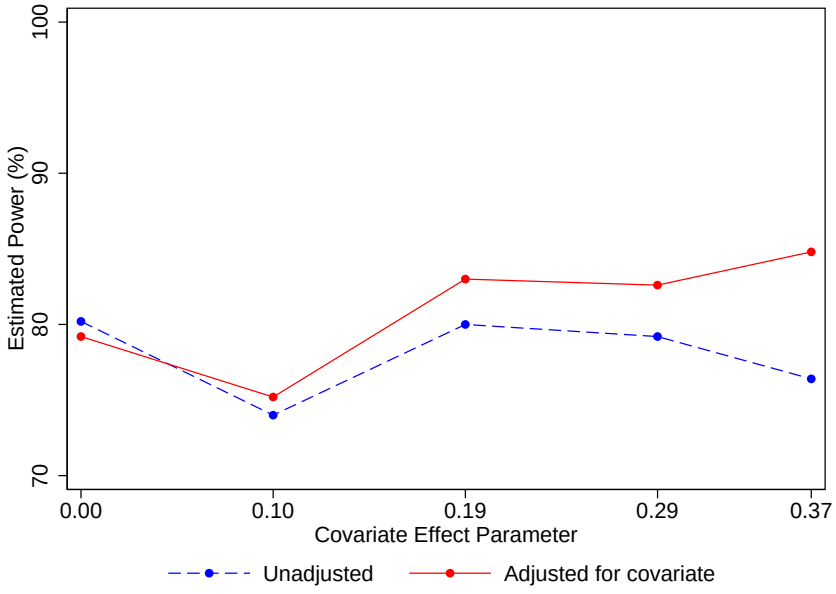


Figure 8.46: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.05. Cluster level covariate. Difference between treatment arms = 0.087. Data from Table 8.22.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.284 (0.009)	0.284 (0.009)	81.6 (1.7)	81.0 (1.8)	0.1
0.0005	0.25 (0.250)	0.291 (0.009)	0.296 (0.009)	79.0 (1.8)	81.0 (1.8)	3.2
0.0005	0.50 (0.500)	0.294 (0.009)	0.328 (0.010)	76.8 (1.9)	79.4 (1.8)	14.8
0.0005	0.75 (0.750)	0.283 (0.009)	0.384 (0.012)	80.8 (1.8)	89.0 (1.4)	52.6
0.0005	0.95 (0.950)	0.287 (0.009)	0.884 (0.028)	77.8 (1.9)	93.4 (1.1)	234.8
0.05	0.00 (0.000)	0.307 (0.010)	0.307 (0.010)	78.2 (1.8)	78.6 (1.8)	0.2
0.05	0.25 (0.256)	0.296 (0.009)	0.302 (0.010)	84.8 (1.6)	84.8 (1.6)	2.9
0.05	0.50 (0.513)	0.287 (0.009)	0.319 (0.010)	77.8 (1.9)	82.8 (1.7)	16.7
0.05	0.75 (0.769)	0.293 (0.009)	0.379 (0.012)	77.4 (1.9)	88.2 (1.4)	50.9
0.05	0.95 (0.975)	0.309 (0.010)	0.855 (0.027)	78.2 (1.8)	97.4 (0.7)	232.5

Continued on next page.

Table 8.22: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Difference between treatment arms = 0.087. (CEF = Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.1	0.00 (0.000)	0.302 (0.010)	0.303 (0.010)	77.0 (1.9)	77.0 (1.9)	0.3
0.1	0.25 (0.264)	0.286 (0.009)	0.293 (0.009)	82.2 (1.7)	82.4 (1.7)	2.9
0.1	0.50 (0.527)	0.283 (0.009)	0.301 (0.010)	80.6 (1.8)	86.4 (1.5)	15.5
0.1	0.75 (0.791)	0.288 (0.009)	0.364 (0.012)	79.6 (1.8)	92.4 (1.2)	51.3
0.1	0.95 (1.001)	0.299 (0.009)	0.869 (0.027)	80.2 (1.8)	97.6 (0.7)	247.4
1	0.00 (0.000)	0.276 (0.009)	0.278 (0.009)	80.2 (1.8)	79.2 (1.8)	−0.1
1	0.25 (0.097)	0.302 (0.010)	0.301 (0.010)	74.0 (2.0)	75.2 (1.9)	0.0
1	0.50 (0.195)	0.283 (0.009)	0.276 (0.009)	80.0 (1.8)	83.0 (1.7)	0.4
1	0.75 (0.292)	0.302 (0.010)	0.296 (0.009)	79.2 (1.8)	82.6 (1.7)	−0.6
1	0.95 (0.370)	0.293 (0.009)	0.267 (0.008)	76.4 (1.9)	84.8 (1.6)	0.9

Table 8.22: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Difference between treatment arms = 0.087. (CEF = Covariate Effect Factor. MCSE=Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.05, and an expected value of the outcome in the control arm of 0.5, are presented in Table 8.23 (page 202). Adjusting for a covariate with ICC of 0.0005 increased estimated power to a maximum of 98% from 83% achieved with the unadjusted model (see Figure 8.47, page). Empirical standard error increased from 0.197 to 0.545, and the estimated realtive asymptotic bias was 257%. Adjusting for a covariate with ICC of 0.05 increased estimated power to a maximum of 100% from 80% achieved with the unadjusted model (see Figure 8.48, page 201). In this case, empirical standard error increased from 0.188 to 0.469, and the estimated relative asymptotic bias was 265%. Adjusting for a cluster level covariate increased estimated power to a maximum of 86% from 83% achieved with the unadjusted model ((see Figure 8.49, page 201)), and empirical standard error decreased from 0.200 to 0.185.

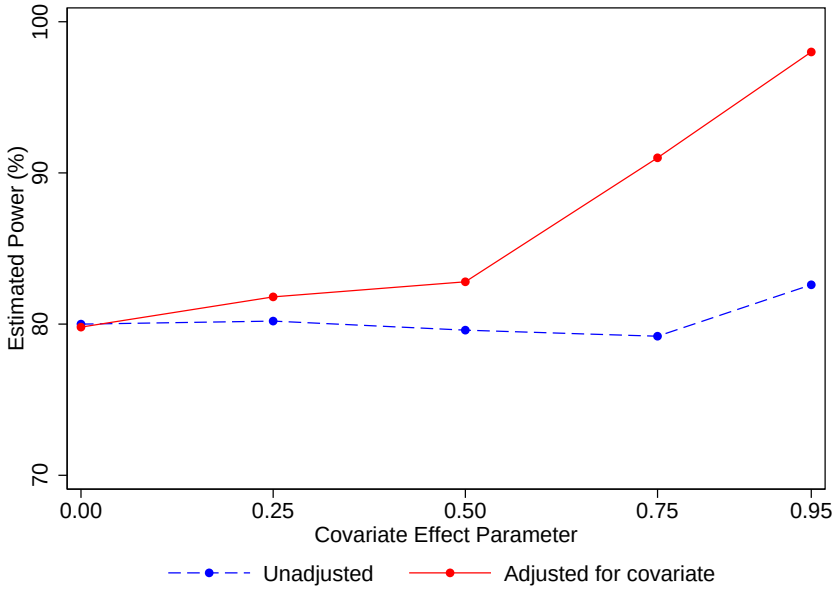


Figure 8.47: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.05. Covariate ICC = 0.0005. Difference between treatment arms = 0.123. Data from Table 8.23.

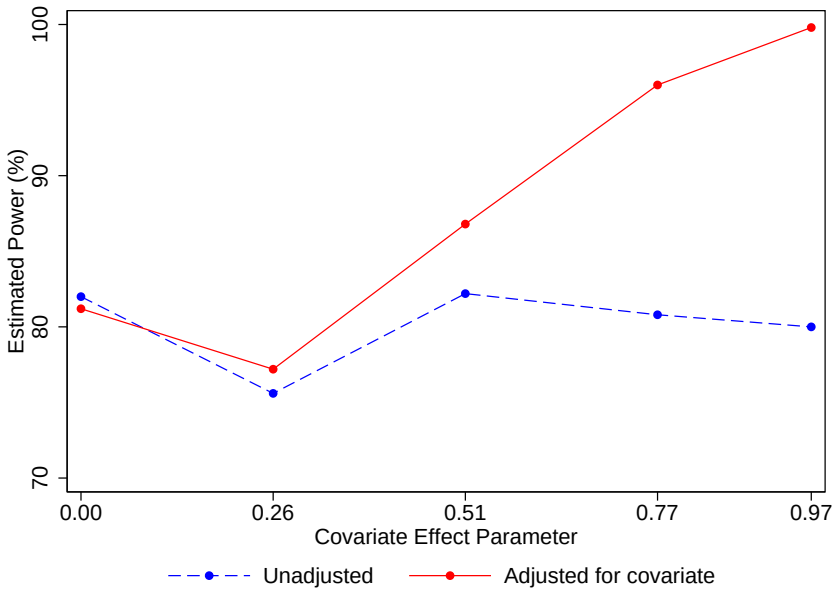


Figure 8.48: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.05. Covariate ICC = 0.05. Difference between treatment arms = 0.123. Data from Table 8.23.

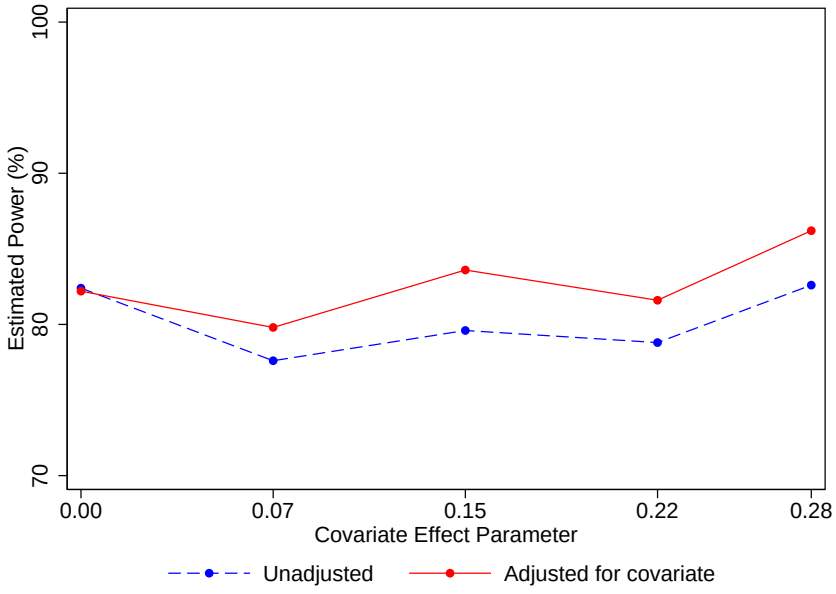


Figure 8.49: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.05. Cluster level covariate. Difference between treatment arms = 0.123. Data from Table 8.23.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.189 (0.006)	0.190 (0.006)	80.0 (1.8)	79.8 (1.8)	0.2
0.0005	0.25 (0.250)	0.182 (0.006)	0.184 (0.006)	80.2 (1.8)	81.8 (1.7)	4.5
0.0005	0.50 (0.500)	0.196 (0.006)	0.220 (0.007)	79.6 (1.8)	82.8 (1.7)	18.9
0.0005	0.75 (0.750)	0.191 (0.006)	0.264 (0.008)	79.2 (1.8)	91.0 (1.3)	61.5
0.0005	0.95 (0.950)	0.197 (0.006)	0.545 (0.017)	82.6 (1.7)	98.0 (0.6)	256.7
0.05	0.00 (0.000)	0.183 (0.006)	0.183 (0.006)	82.0 (1.7)	81.2 (1.7)	0.1
0.05	0.25 (0.256)	0.193 (0.006)	0.198 (0.006)	75.6 (1.9)	77.2 (1.9)	4.6
0.05	0.50 (0.513)	0.189 (0.006)	0.204 (0.006)	82.2 (1.7)	86.8 (1.5)	19.2
0.05	0.75 (0.769)	0.197 (0.006)	0.248 (0.008)	80.8 (1.8)	96.0 (0.9)	60.9
0.05	0.95 (0.975)	0.188 (0.006)	0.469 (0.015)	80.0 (1.8)	99.8 (0.2)	265.4

Continued on next page.

Table 8.23: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Difference between treatment arms = 0.123. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Continued from previous page.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.1	0.00 (0.000)	0.195 (0.006)	0.196 (0.006)	79.8 (1.8)	79.8 (1.8)	0.1
0.1	0.25 (0.231)	0.191 (0.006)	0.195 (0.006)	78.0 (1.9)	79.8 (1.8)	3.2
0.1	0.50 (0.461)	0.190 (0.006)	0.201 (0.006)	75.0 (1.9)	83.8 (1.6)	14.1
0.1	0.75 (0.692)	0.191 (0.006)	0.217 (0.007)	79.2 (1.8)	93.2 (1.1)	40.5
0.1	0.95 (0.877)	0.193 (0.006)	0.273 (0.009)	80.8 (1.8)	98.4 (0.6)	96.6
1	0.00 (0.000)	0.187 (0.006)	0.189 (0.006)	82.4 (1.7)	82.2 (1.7)	0.1
1	0.25 (0.073)	0.198 (0.006)	0.196 (0.006)	77.6 (1.9)	79.8 (1.8)	0.2
1	0.50 (0.146)	0.187 (0.006)	0.183 (0.006)	79.6 (1.8)	83.6 (1.7)	0.0
1	0.75 (0.219)	0.189 (0.006)	0.180 (0.006)	78.8 (1.8)	81.6 (1.7)	0.0
1	0.95 (0.277)	0.200 (0.006)	0.185 (0.006)	82.6 (1.7)	86.2 (1.5)	1.5

Table 8.23: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Difference between treatment arms = 0.123. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.1, and an expected value of 0.1 in the control treatment arm, are presented in Table 8.24 (page 206). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 88% from 74% (see Figure 8.50), while empirical standard error increased from 0.353 to 1.118. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 95% from 80% (see Figure 8.51, page 205). Empirical standard error increased from 0.326 to 1.037, and estimated relative asymptotic bias was 242%. Adjusting for a cluster level covariate increased estimated power to a maximum of 92% from 80% (see Figure 8.52, page 205). Here, empirical standard error reduced from 0.341 to 0.280.

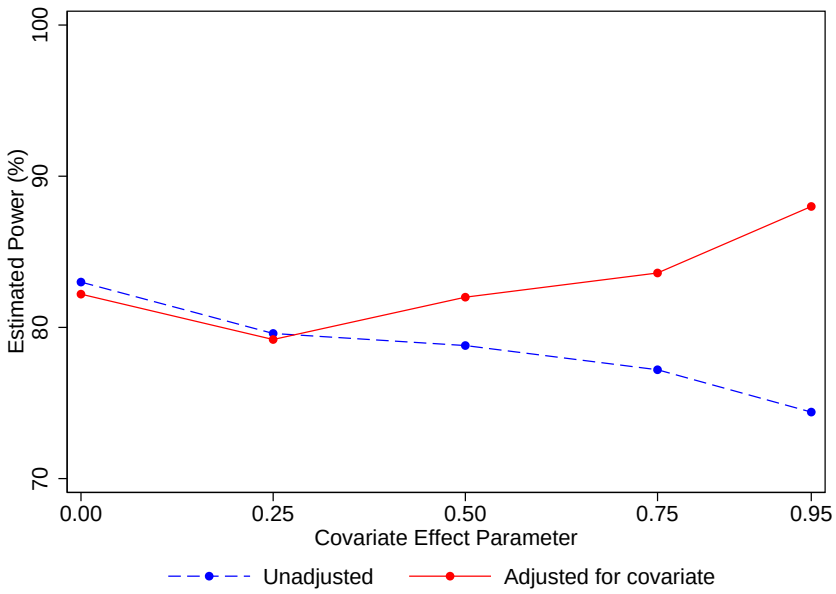


Figure 8.50: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.1. Covariate ICC = 0.0005. Difference between treatment arms = 0.095. Data from Table 8.24.

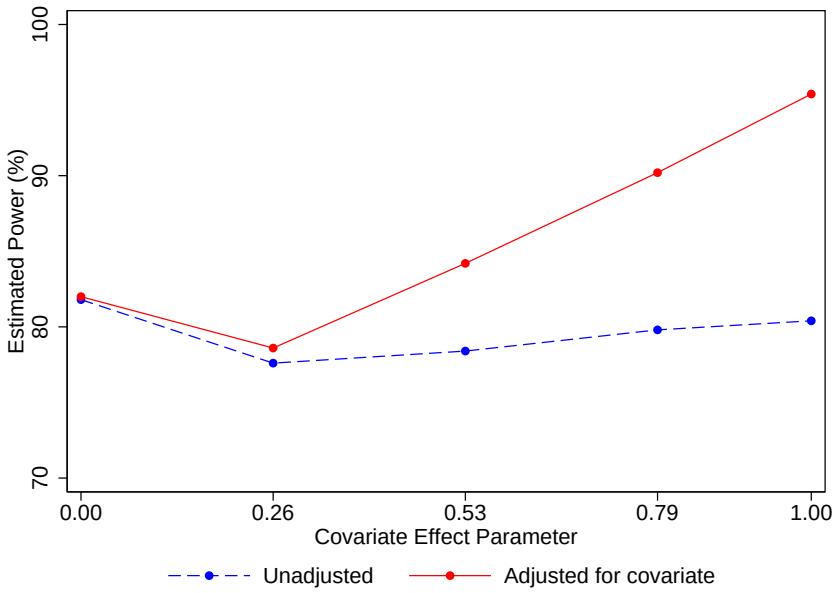


Figure 8.51: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.1. Covariate ICC = 0.1. Difference between treatment arms = 0.095. Data from Table 8.24.

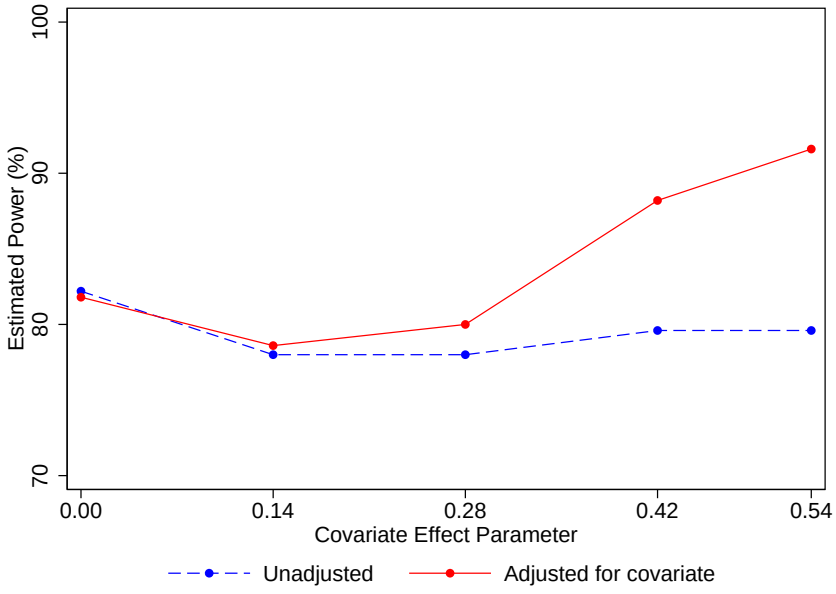


Figure 8.52: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.1. Cluster level covariate. Difference between treatment arms = 0.095. Data from Table 8.24.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.330 (0.010)	0.332 (0.011)	83.0 (1.7)	82.2 (1.7)	0.4
0.0005	0.25 (0.250)	0.315 (0.010)	0.322 (0.010)	79.6 (1.8)	79.2 (1.8)	3.1
0.0005	0.50 (0.500)	0.329 (0.010)	0.367 (0.012)	78.8 (1.8)	82.0 (1.7)	16.2
0.0005	0.75 (0.750)	0.330 (0.010)	0.483 (0.015)	77.2 (1.9)	83.6 (1.7)	53.9
0.0005	0.95 (0.950)	0.353 (0.011)	1.118 (0.035)	74.4 (2.0)	88.0 (1.5)	239.1
0.1	0.00 (0.000)	0.310 (0.010)	0.311 (0.010)	81.8 (1.7)	82.0 (1.7)	0.2
0.1	0.25 (0.264)	0.355 (0.011)	0.365 (0.012)	77.6 (1.9)	78.6 (1.8)	3.9
0.1	0.50 (0.527)	0.338 (0.011)	0.377 (0.012)	78.4 (1.8)	84.2 (1.6)	17.6
0.1	0.75 (0.791)	0.323 (0.010)	0.418 (0.013)	79.8 (1.8)	90.2 (1.3)	51.2
0.1	0.95 (1.001)	0.326 (0.010)	1.037 (0.033)	80.4 (1.8)	95.4 (0.9)	241.7
1	0.00 (0.000)	0.348 (0.011)	0.352 (0.011)	82.2 (1.7)	81.8 (1.7)	−0.1
1	0.25 (0.141)	0.336 (0.011)	0.331 (0.010)	78.0 (1.9)	78.6 (1.8)	−0.1
1	0.50 (0.283)	0.322 (0.010)	0.315 (0.010)	78.0 (1.9)	80.0 (1.8)	0.0
1	0.75 (0.424)	0.334 (0.011)	0.295 (0.009)	79.6 (1.8)	88.2 (1.4)	−0.3
1	0.95 (0.537)	0.341 (0.011)	0.280 (0.009)	79.6 (1.8)	91.6 (1.2)	1.5

Table 8.24: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.1 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. Difference between treatment arms = 0.095. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Results for an outcome with an ICC of 0.1, and an expected value of 0.5 in the control treatment arm, are presented in Table 8.25 (page 209). Adjusting for a covariate with an ICC of 0.0005 increased estimated power to a maximum of 96% from 80% (see Figure 8.53), while empirical standard error increased from 0.215 to 0.692. Adjusting for a covariate with an ICC of 0.1 increased estimated power to a maximum of 99% from 80% (see Figure 8.54, page 208). Empirical standard error increased from 0.210 to 0.583, and estimated relative asymptotic bias was 268%. Adjusting for a cluster level covariate increased estimated power to a maximum of 94% from 83% (see Figure 8.55, page 208). Here, empirical standard error reduced from 0.212 to 0.180.

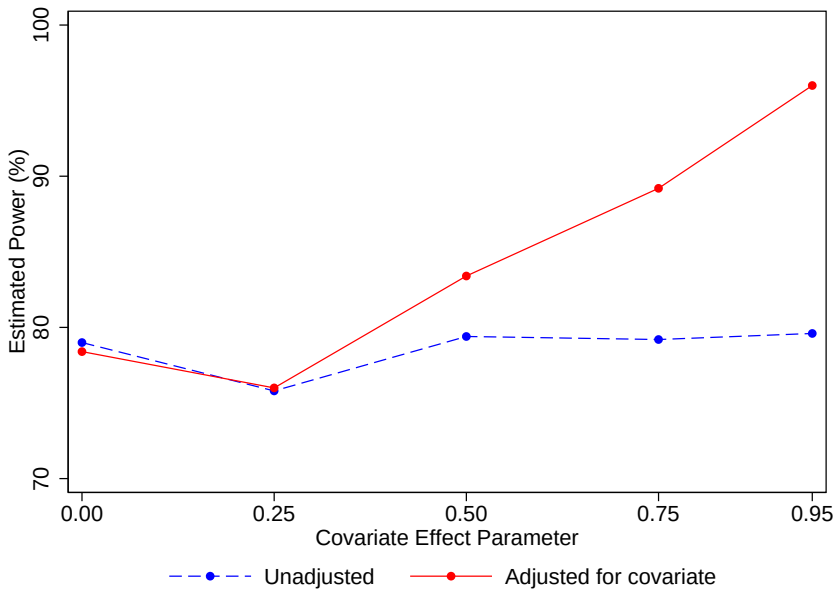


Figure 8.53: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.1. Covariate ICC = 0.0005. Difference between treatment arms = 0.133. Data from Table 8.25.

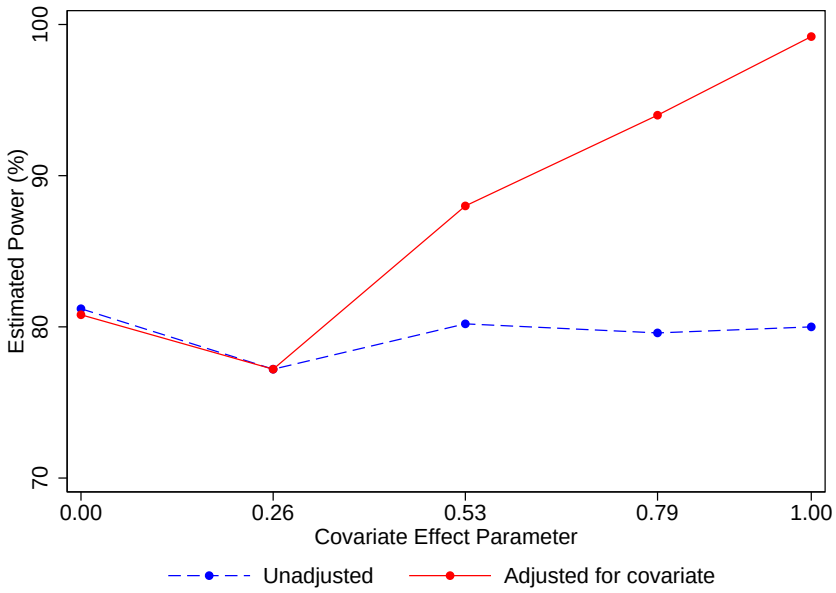


Figure 8.54: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.5. Outcome ICC = 0.1. Covariate ICC = 0.1. Difference between treatment arms = 0.133. Data from Table 8.25.

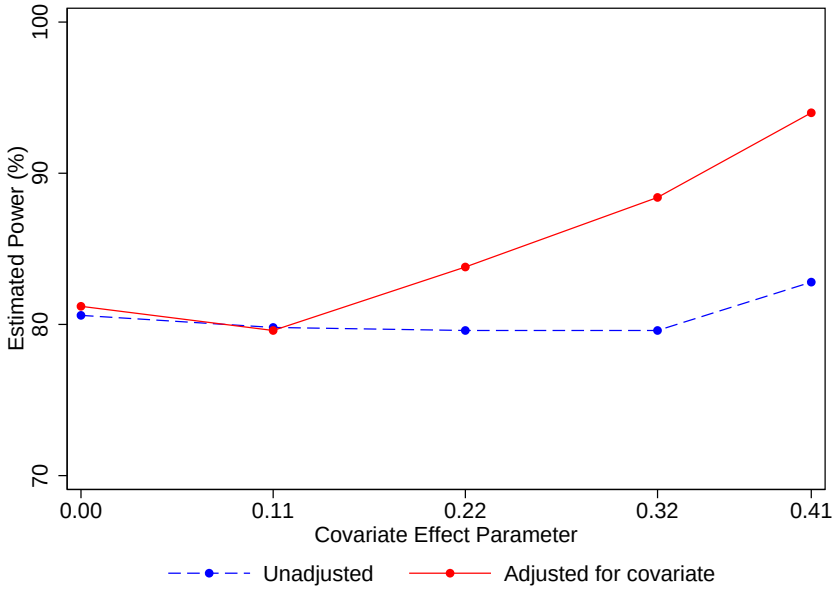


Figure 8.55: Estimated power when using a logistic mixed effects model, unadjusted and adjusted for a covariate. 60 clusters per treatment arm, cluster size 5. Expected value of outcome in control arm = 0.1. Outcome ICC = 0.5. Cluster level covariate. Difference between treatment arms = 0.133. Data from Table 8.25.

		Empirical SE (MCSE)		Estimated Power % (MCSE)		Estimated Relative
Covariate ICC	CEF (Parameter)	Unadjusted	Adjusted	Unadjusted	Adjusted	Asymptotic Bias %
0.0005	0.00 (0.000)	0.214 (0.007)	0.214 (0.007)	79.0 (1.8)	78.4 (1.8)	0.1
0.0005	0.25 (0.250)	0.225 (0.007)	0.231 (0.007)	75.8 (1.9)	76.0 (1.9)	3.4
0.0005	0.50 (0.500)	0.224 (0.007)	0.260 (0.008)	79.4 (1.8)	83.4 (1.7)	19.2
0.0005	0.75 (0.750)	0.222 (0.007)	0.327 (0.010)	79.2 (1.8)	89.2 (1.4)	61.1
0.0005	0.95 (0.950)	0.215 (0.007)	0.692 (0.022)	79.6 (1.8)	96.0 (0.9)	268.1
0.1	0.00 (0.000)	0.216 (0.007)	0.217 (0.007)	81.2 (1.7)	80.8 (1.8)	0.3
0.1	0.25 (0.264)	0.233 (0.007)	0.240 (0.008)	77.2 (1.9)	77.2 (1.9)	4.4
0.1	0.50 (0.527)	0.213 (0.007)	0.237 (0.007)	80.2 (1.8)	88.0 (1.5)	18.4
0.1	0.75 (0.791)	0.222 (0.007)	0.294 (0.009)	79.6 (1.8)	94.0 (1.1)	61.3
0.1	0.95 (1.001)	0.210 (0.007)	0.583 (0.018)	80.0 (1.8)	99.2 (0.4)	268.2
1	0.00 (0.000)	0.219 (0.007)	0.219 (0.007)	80.6 (1.8)	81.2 (1.7)	0.0
1	0.25 (0.108)	0.203 (0.006)	0.202 (0.006)	79.8 (1.8)	79.6 (1.8)	−0.4
1	0.50 (0.215)	0.219 (0.007)	0.209 (0.007)	79.6 (1.8)	83.8 (1.6)	0.5
1	0.75 (0.323)	0.222 (0.007)	0.196 (0.006)	79.6 (1.8)	88.4 (1.4)	−0.3
1	0.95 (0.409)	0.212 (0.007)	0.180 (0.006)	82.8 (1.7)	94.0 (1.1)	1.8

Table 8.25: Empirical standard error, estimated power, and asymptotic bias, when using a logistic mixed effects model for an outcome with expected value of 0.5 in the control treatment arm. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. Difference between treatment arms = 0.133. (CEF = Covariate Effect Factor. MCSE = Monte Carlo Standard Error.)

Overall, with a cluster size of five, adjusting for a covariate with the same ICC as the outcome increased estimated power most notably. However, with a cluster size of 30 and an outcome with an ICC of 0.05 or 0.1 adjusting for a cluster level covariate often increased power as much as or more than adjusting for a covariate with an ICC of 0.05 or 0.1. For an outcome with an ICC of 0.0005, adjusting for a covariate with an ICC of 0.05 or a cluster level covariate, did not significantly increased estimated power. With a cluster size of five and an outcome with an ICC of 0.05, adjusting for a covariate with a very small ICC increased estimated power more than adjusting for a cluster level covariate. But with a cluster size of 30, adjusting for a cluster level covariate increased estimated power more than adjusting for a covariate with an ICC of 0.0005.

8.5 Models using separate within-cluster and contextual covariate effect parameters

Tables 8.26 to 8.31 give estimated power when using a linear mixed effects model and adjusting for a covariate using a single covariate effect parameter, or separate within-cluster and contextual covariate effect parameters.

Tables 8.26 to 8.28 present results for simulated CRTs with a cluster size of 30 and 20 clusters in each treatment arm.

For an outcome with an ICC of 0.0005 (Table 8.26, page 212) estimated power was not significantly greater using separate covariate effect parameters for any scenarios simulated. In fact, for all scenarios estimated power was slightly reduced when using separate covariate effect parameters.

			Estimated Power % (MCSE)	
Covariate ICC	CEF	RCE (γ_w, γ_b)	Adjusted	Separate Effects
0.0005	0.75	-0.67 (0.335, -0.224)	82.3 (0.5)	81.0 (0.6)
0.0005	0.95	-0.67 (0.425, -0.283)	86.0 (0.5)	84.7 (0.5)
0.05	0.75	-0.67 (0.034, -0.022)	77.4 (0.6)	76.3 (0.6)
0.05	0.95	-0.67 (0.042, -0.028)	77.6 (0.6)	76.6 (0.6)
0.1	0.75	-0.67 (0.024, -0.016)	77.7 (0.6)	76.9 (0.6)
0.1	0.95	-0.67 (0.030, -0.020)	76.6 (0.6)	75.8 (0.6)
0.0005	0.75	-0.50 (0.335, -0.168)	82.5 (0.5)	81.7 (0.5)
0.0005	0.95	-0.50 (0.425, -0.212)	85.2 (0.5)	84.6 (0.5)
0.05	0.75	-0.50 (0.034, -0.017)	77.2 (0.6)	76.8 (0.6)
0.05	0.95	-0.50 (0.042, -0.021)	77.4 (0.6)	76.2 (0.6)
0.1	0.75	-0.50 (0.024, -0.012)	77.2 (0.6)	76.2 (0.6)
0.1	0.95	-0.50 (0.030, -0.015)	77.7 (0.6)	76.8 (0.6)
0.0005	0.75	0.00 (0.335, 0.000)	82.2 (0.5)	81.1 (0.6)
0.0005	0.95	0.00 (0.425, 0.000)	85.5 (0.5)	84.3 (0.5)
0.05	0.75	0.00 (0.034, 0.000)	76.8 (0.6)	75.9 (0.6)
0.05	0.95	0.00 (0.042, 0.000)	78.3 (0.6)	77.2 (0.6)
0.1	0.75	0.00 (0.024, 0.000)	78.4 (0.6)	77.1 (0.6)
0.1	0.95	0.00 (0.030, 0.000)	77.2 (0.6)	76.4 (0.6)
0.0005	0.75	1.00 (0.335, 0.335)	82.7 (0.5)	81.7 (0.5)
0.0005	0.95	1.00 (0.425, 0.425)	85.3 (0.5)	84.2 (0.5)
0.05	0.75	1.00 (0.034, 0.034)	76.5 (0.6)	75.4 (0.6)
0.05	0.95	1.00 (0.042, 0.042)	78.0 (0.6)	77.4 (0.6)
0.1	0.75	1.00 (0.024, 0.024)	77.1 (0.6)	75.9 (0.6)
0.1	0.95	1.00 (0.030, 0.030)	77.0 (0.6)	75.9 (0.6)
0.0005	0.75	2.00 (0.335, 0.671)	82.2 (0.5)	81.0 (0.6)
0.0005	0.95	2.00 (0.425, 0.850)	85.1 (0.5)	84.1 (0.5)
0.05	0.75	2.00 (0.034, 0.067)	77.6 (0.6)	76.8 (0.6)
0.05	0.95	2.00 (0.042, 0.085)	77.6 (0.6)	76.9 (0.6)
0.1	0.75	2.00 (0.024, 0.047)	77.7 (0.6)	76.8 (0.6)
0.1	0.95	2.00 (0.030, 0.060)	77.7 (0.6)	77.0 (0.6)

Table 8.26: Estimated power with a linear mixed effects model using a single covariate effect parameter (Adjusted) or separate covariate effects (Separate Effects). 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.0005. (CEF = Covariate Effect Factor. RCE = Ratio of Covariate Effects. MCSE = Monte Carlo Standard Error.)

For an outcome with an ICC of 0.05 (Table 8.27, page 215), estimated power was greater using separate covariate effect parameters when the covariate had an ICC of 0.05 or 0.1, and the between-cluster covariate effect parameter was equal to or twice the size of the within-cluster covariate effect parameter. Otherwise, estimated power was reduced when using separate covariate effect parameters. The greatest increase in power was from 87% to 97% for a covariate with an ICC of 0.1 and when the between-cluster covariate effect was twice the magnitude of the within-cluster covariate effect, as shown in Figure 8.57 (page 214).

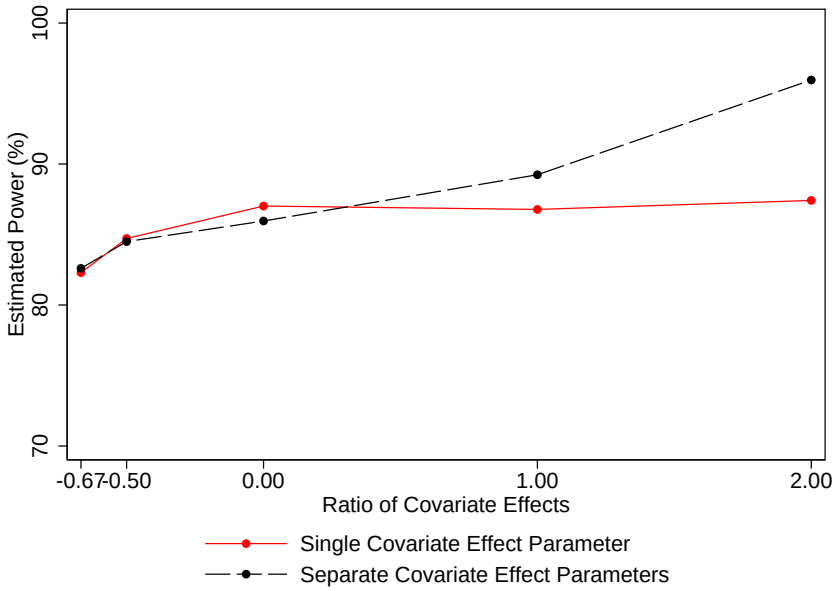


Figure 8.56: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 10 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Covariate ICC = 0.05. Data from Table 8.27.

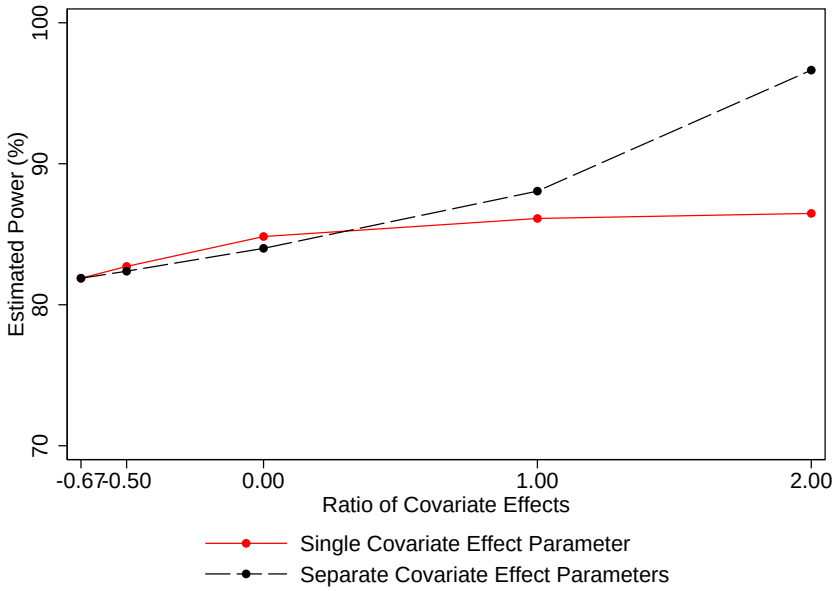


Figure 8.57: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 10 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. Covariate ICC = 0.1. Data from Table 8.27.

			Estimated Power % (MCSE)	
Covariate ICC	CEF	RCE (γ_w, γ_b)	Adjusted	Separate Effects
0.0005	0.75	-0.67 (0.731, -0.487)	87.6 (0.5)	86.8 (0.5)
0.0005	0.95	-0.67 (0.926, -0.617)	93.4 (0.4)	92.7 (0.4)
0.05	0.75	-0.67 (0.335, -0.224)	81.8 (0.5)	81.6 (0.5)
0.05	0.95	-0.67 (0.425, -0.283)	82.3 (0.5)	82.6 (0.5)
0.1	0.75	-0.67 (0.237, -0.158)	80.8 (0.6)	80.7 (0.6)
0.1	0.95	-0.67 (0.300, -0.200)	81.9 (0.5)	81.9 (0.5)
0.0005	0.75	-0.50 (0.731, -0.366)	88.2 (0.5)	87.5 (0.5)
0.0005	0.95	-0.50 (0.926, -0.463)	92.9 (0.4)	92.0 (0.4)
0.05	0.75	-0.50 (0.335, -0.168)	83.1 (0.5)	82.6 (0.5)
0.05	0.95	-0.50 (0.425, -0.212)	84.7 (0.5)	84.5 (0.5)
0.1	0.75	-0.50 (0.237, -0.119)	82.4 (0.5)	81.9 (0.5)
0.1	0.95	-0.50 (0.300, -0.150)	82.7 (0.5)	82.4 (0.5)
0.0005	0.75	0.00 (0.731, 0.000)	88.5 (0.5)	87.2 (0.5)
0.0005	0.95	0.00 (0.926, 0.000)	93.4 (0.4)	92.7 (0.4)
0.05	0.75	0.00 (0.335, 0.000)	83.7 (0.5)	82.9 (0.5)
0.05	0.95	0.00 (0.425, 0.000)	87.0 (0.5)	86.0 (0.5)
0.1	0.75	0.00 (0.237, 0.000)	83.3 (0.5)	82.2 (0.5)
0.1	0.95	0.00 (0.300, 0.000)	84.8 (0.5)	84.0 (0.5)
0.0005	0.75	1.00 (0.731, 0.731)	90.0 (0.4)	89.2 (0.4)
0.0005	0.95	1.00 (0.926, 0.926)	93.5 (0.3)	92.8 (0.4)
0.05	0.75	1.00 (0.335, 0.335)	83.8 (0.5)	84.9 (0.5)
0.05	0.95	1.00 (0.425, 0.425)	86.8 (0.5)	89.2 (0.4)
0.1	0.75	1.00 (0.237, 0.237)	82.7 (0.5)	83.2 (0.5)
0.1	0.95	1.00 (0.300, 0.300)	86.1 (0.5)	88.1 (0.5)
0.0005	0.75	2.00 (0.731, 1.462)	88.3 (0.5)	87.5 (0.5)
0.0005	0.95	2.00 (0.926, 1.852)	93.7 (0.3)	93.2 (0.4)
0.05	0.75	2.00 (0.335, 0.671)	84.3 (0.5)	89.9 (0.4)
0.05	0.95	2.00 (0.425, 0.850)	87.4 (0.5)	96.0 (0.3)
0.1	0.75	2.00 (0.237, 0.474)	84.3 (0.5)	90.6 (0.4)
0.1	0.95	2.00 (0.300, 0.601)	86.5 (0.5)	96.6 (0.3)

Table 8.27: Estimated Power with linear mixed effects model using a single covariate effect parameter (Adjusted) and separate covariate effects models (Separate effects). 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.05. (CEF = Covariate Effect Factor. RCE = Ratio of Covariate Effects. MCSE = Monte Carlo Standard Error.)

For an outcome with an ICC of 0.1 (Table 8.28, page 218), estimated power was greater using separate covariate effect parameters when the covariate had an ICC of 0.05 or 0.1, and the between-cluster covariate effect parameter was equal to or twice the size of the within-cluster covariate effect parameter. In some other cases, power was marginally higher when using separate covariate effect parameters. The greatest increase in power was from 88% to 96% for a covariate with an ICC of 0.1 and when the between-cluster covariate effect was twice the magnitude of the within-cluster covariate effect.

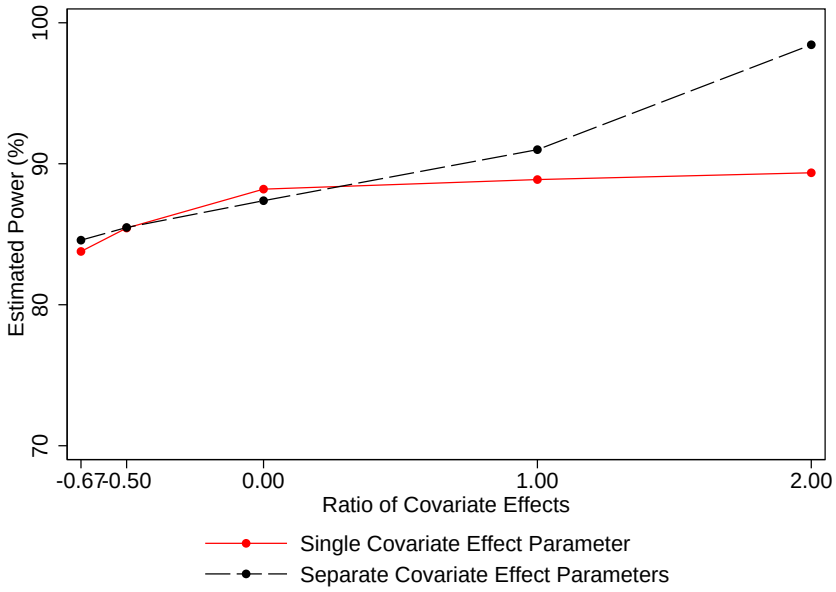


Figure 8.58: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 10 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. Covariate ICC = 0.05. Data from Table 8.28.

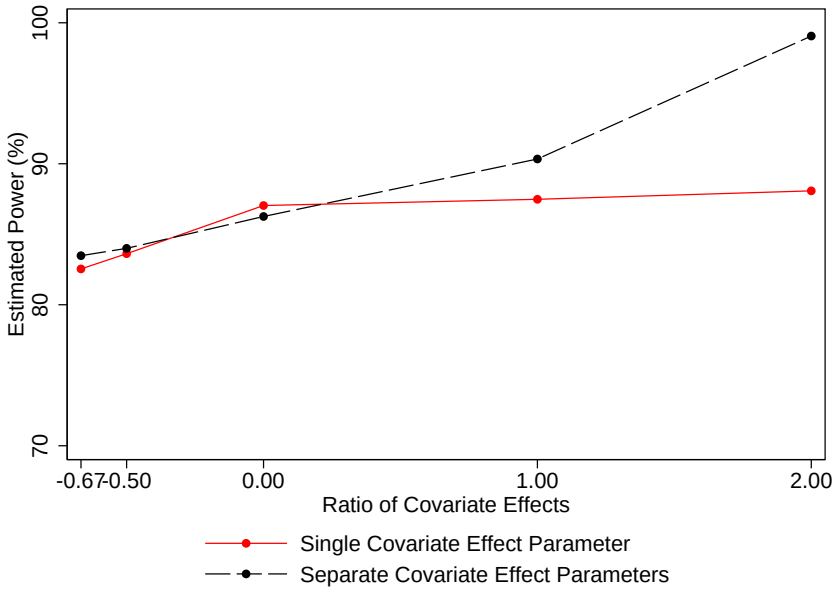


Figure 8.59: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 10 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. Covariate ICC = 0.1. Data from Table 8.28.

			Estimated Power % (MCSE)	
Covariate ICC	CEF	RCE (γ_w, γ_b)	Adjusted	Separate Effects
0.0005	0.75	-0.67 (0.712, -0.474)	85.2 (0.5)	84.4 (0.5)
0.0005	0.95	-0.67 (0.901, -0.601)	88.4 (0.5)	87.8 (0.5)
0.05	0.75	-0.67 (0.474, -0.316)	82.6 (0.5)	82.4 (0.5)
0.05	0.95	-0.67 (0.601, -0.401)	83.8 (0.5)	84.6 (0.5)
0.1	0.75	-0.67 (0.335, -0.224)	81.8 (0.5)	81.5 (0.5)
0.1	0.95	-0.67 (0.425, -0.283)	82.5 (0.5)	83.5 (0.5)
0.0005	0.75	-0.50 (0.712, -0.356)	84.3 (0.5)	83.7 (0.5)
0.0005	0.95	-0.50 (0.901, -0.451)	88.2 (0.5)	87.5 (0.5)
0.05	0.75	-0.50 (0.474, -0.237)	84.4 (0.5)	83.9 (0.5)
0.05	0.95	-0.50 (0.601, -0.300)	85.4 (0.5)	85.5 (0.5)
0.1	0.75	-0.50 (0.335, -0.168)	82.6 (0.5)	82.0 (0.5)
0.1	0.95	-0.50 (0.425, -0.212)	83.6 (0.5)	84.0 (0.5)
0.0005	0.75	0.00 (0.712, 0.000)	85.2 (0.5)	84.4 (0.5)
0.0005	0.95	0.00 (0.901, 0.000)	88.4 (0.5)	87.4 (0.5)
0.05	0.75	0.00 (0.474, 0.000)	85.3 (0.5)	84.3 (0.5)
0.05	0.95	0.00 (0.601, 0.000)	88.2 (0.5)	87.4 (0.5)
0.1	0.75	0.00 (0.335, 0.000)	84.0 (0.5)	83.2 (0.5)
0.1	0.95	0.00 (0.425, 0.000)	87.0 (0.5)	86.3 (0.5)
0.0005	0.75	1.00 (0.712, 0.712)	85.9 (0.5)	85.2 (0.5)
0.0005	0.95	1.00 (0.901, 0.901)	87.6 (0.5)	86.8 (0.5)
0.05	0.75	1.00 (0.474, 0.474)	85.4 (0.5)	86.6 (0.5)
0.05	0.95	1.00 (0.601, 0.601)	88.9 (0.4)	91.0 (0.4)
0.1	0.75	1.00 (0.335, 0.335)	84.5 (0.5)	85.7 (0.5)
0.1	0.95	1.00 (0.425, 0.425)	87.5 (0.5)	90.3 (0.4)
0.0005	0.75	2.00 (0.712, 1.423)	86.4 (0.5)	85.6 (0.5)
0.0005	0.95	2.00 (0.901, 1.803)	87.5 (0.5)	86.4 (0.5)
0.05	0.75	2.00 (0.474, 0.949)	85.8 (0.5)	92.5 (0.4)
0.05	0.95	2.00 (0.601, 1.202)	89.4 (0.4)	98.4 (0.2)
0.1	0.75	2.00 (0.335, 0.671)	83.9 (0.5)	93.1 (0.4)
0.1	0.95	2.00 (0.425, 0.850)	88.1 (0.5)	99.1 (0.1)

Table 8.28: Estimated Power with a linear mixed effects model using a single covariate effect parameter (Adjusted) or separate covariate effects models (Separate effects). 20 clusters per treatment arm, cluster size 30. Outcome ICC = 0.1. (CEF = Covariate Effect Factor. RCE = Ratio of Covariate Effects. MCSE = Monte Carlo Standard Error.)

Tables 8.29 to 8.31 present results for simulated CRTs with a cluster size of 5 and 60 clusters in each treatment arm.

For an outcome with an ICC of 0.0005 (Table 8.29, page 220), estimated power was marginally smaller in all scenarios when using a linear mixed effects model with separate within-cluster and contextual covariate effect parameters.

			Estimated Power % (MCSE)	
Covariate ICC	CEF	RCE (γ_w, γ_b)	Adjusted	Separate Effects
0.0005	0.75	-0.67 (0.335, -0.224)	83.0 (0.5)	82.8 (0.5)
0.0005	0.95	-0.67 (0.425, -0.283)	85.8 (0.5)	85.7 (0.5)
0.05	0.75	-0.67 (0.034, -0.022)	78.3 (0.6)	78.4 (0.6)
0.05	0.95	-0.67 (0.042, -0.028)	79.3 (0.6)	79.2 (0.6)
0.1	0.75	-0.67 (0.024, -0.016)	78.8 (0.6)	78.4 (0.6)
0.1	0.95	-0.67 (0.030, -0.020)	77.7 (0.6)	77.6 (0.6)
0.0005	0.75	-0.50 (0.335, -0.168)	82.7 (0.5)	82.6 (0.5)
0.0005	0.95	-0.50 (0.425, -0.212)	85.1 (0.5)	84.8 (0.5)
0.05	0.75	-0.50 (0.034, -0.017)	78.1 (0.6)	77.8 (0.6)
0.05	0.95	-0.50 (0.042, -0.021)	78.9 (0.6)	78.3 (0.6)
0.1	0.75	-0.50 (0.024, -0.012)	80.1 (0.6)	79.9 (0.6)
0.1	0.95	-0.50 (0.030, -0.015)	78.5 (0.6)	78.4 (0.6)
0.0005	0.75	0.00 (0.335, 0.000)	83.7 (0.5)	83.4 (0.5)
0.0005	0.95	0.00 (0.425, 0.000)	86.0 (0.5)	85.6 (0.5)
0.05	0.75	0.00 (0.034, 0.000)	79.4 (0.6)	78.9 (0.6)
0.05	0.95	0.00 (0.042, 0.000)	79.5 (0.6)	79.1 (0.6)
0.1	0.75	0.00 (0.024, 0.000)	78.5 (0.6)	78.4 (0.6)
0.1	0.95	0.00 (0.030, 0.000)	78.2 (0.6)	77.9 (0.6)
0.0005	0.75	1.00 (0.335, 0.335)	83.8 (0.5)	83.5 (0.5)
0.0005	0.95	1.00 (0.425, 0.425)	86.0 (0.5)	85.7 (0.5)
0.05	0.75	1.00 (0.034, 0.034)	78.8 (0.6)	78.6 (0.6)
0.05	0.95	1.00 (0.042, 0.042)	78.7 (0.6)	78.5 (0.6)
0.1	0.75	1.00 (0.024, 0.024)	78.4 (0.6)	78.6 (0.6)
0.1	0.95	1.00 (0.030, 0.030)	77.9 (0.6)	77.7 (0.6)
0.0005	0.75	2.00 (0.335, 0.671)	83.8 (0.5)	83.3 (0.5)
0.0005	0.95	2.00 (0.425, 0.850)	85.9 (0.5)	85.6 (0.5)
0.05	0.75	2.00 (0.034, 0.067)	78.0 (0.6)	78.0 (0.6)
0.05	0.95	2.00 (0.042, 0.085)	78.6 (0.6)	78.8 (0.6)
0.1	0.75	2.00 (0.024, 0.047)	78.4 (0.6)	78.3 (0.6)
0.1	0.95	2.00 (0.030, 0.060)	78.8 (0.6)	78.6 (0.6)

Table 8.29: Estimated Power with a linear mixed effects model using a single covariate effect parameter (Adjusted) or separate covariate effects (Separate Effects). 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.0005. (CEF=Covariate Effect Factor. RCE=Ratio of Covariate Effects. MCSE=Monte Carlo Standard Error.)

For an outcome with an ICC of 0.05 or 0.1 (Tables 8.30 and 8.31, pages 224 and 225), estimated power was marginally greater when using separate covariate effect parameters when the covariate had an ICC of 0.05 or 0.1, and the between-cluster covariate effect parameter was twice the magnitude of the within-cluster covariate effect parameter. Otherwise, estimated power was similar or marginally reduced using separate covariate effect parameters.

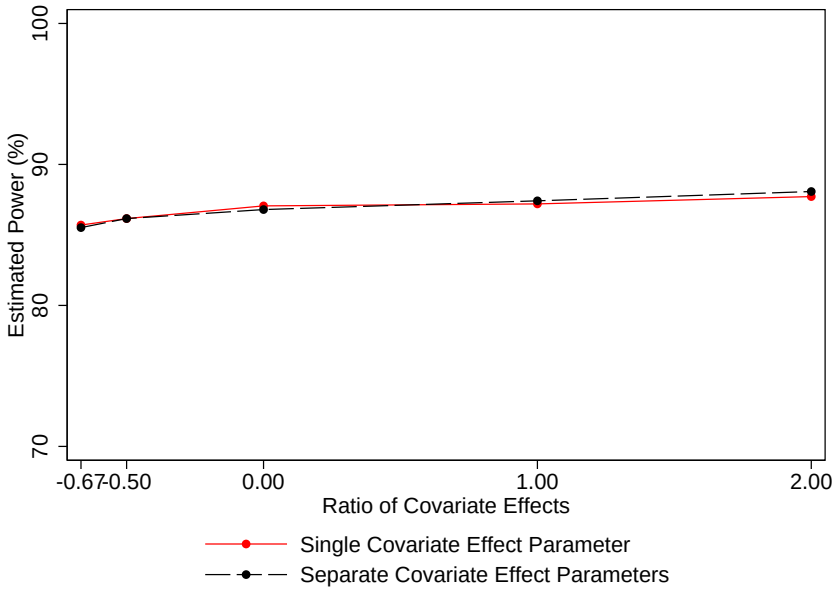


Figure 8.60: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Covariate ICC = 0.05. Data from Table 8.30.

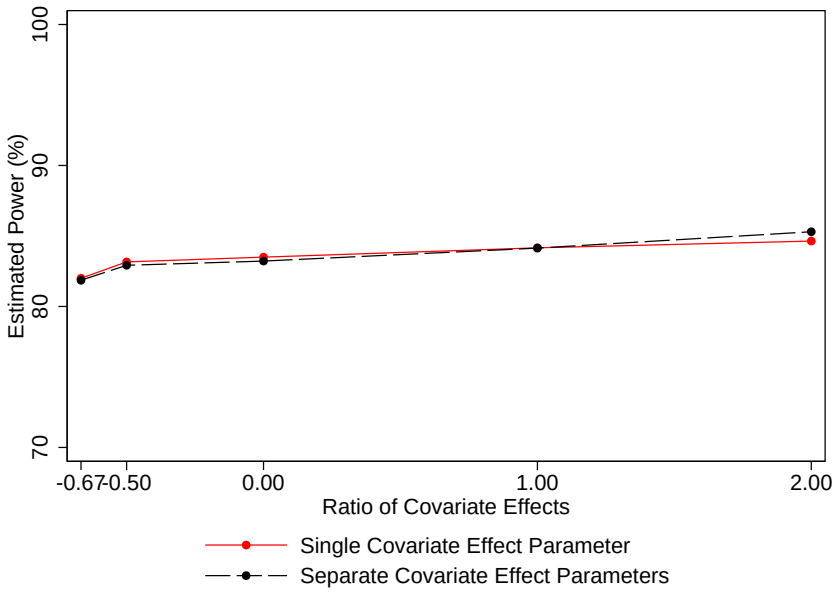


Figure 8.61: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. Covariate ICC = 0.1. Data from Table 8.30.

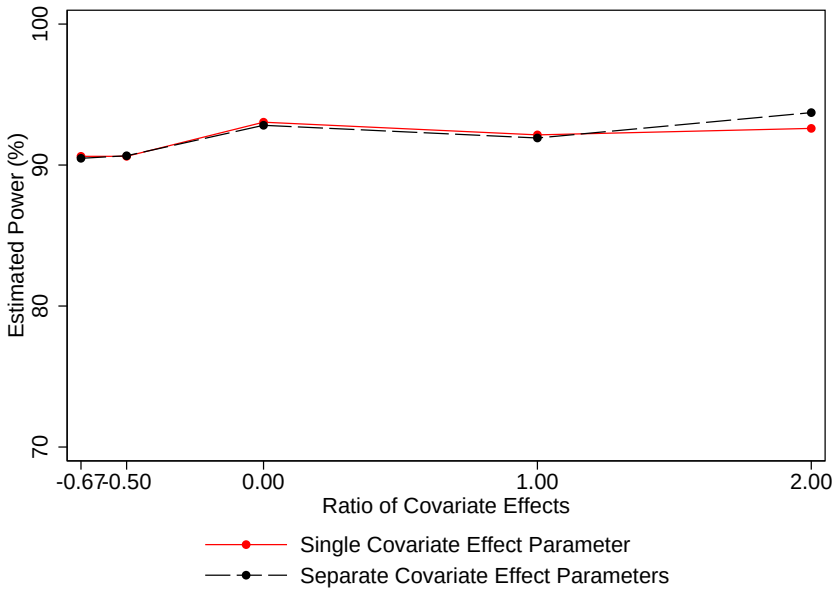


Figure 8.62: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. Covariate ICC = 0.05. Data from Table 8.31.

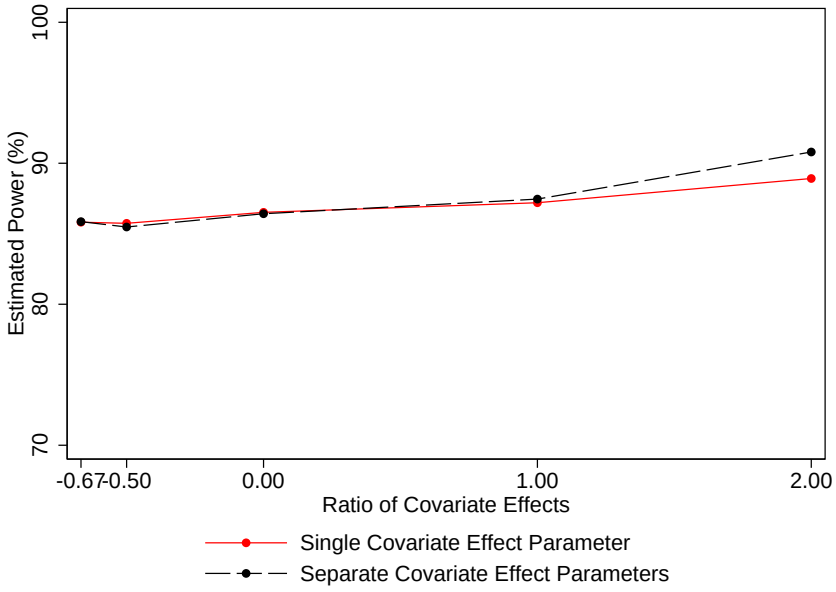


Figure 8.63: Estimated power with linear mixed effects models using a single covariate effect parameter or separate covariate effect parameters. 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. Covariate ICC = 0.1. Data from Table 8.31.

			Estimated Power % (MCSE)	
Covariate ICC	CEF	RCE (γ_w, γ_b)	Adjusted	Separate effects
0.0005	0.75	-0.67 (0.731, -0.487)	96.9 (0.2)	96.8 (0.2)
0.0005	0.95	-0.67 (0.926, -0.617)	99.9 (<0.1)	99.9 (<0.1)
0.05	0.75	-0.67 (0.335, -0.224)	83.8 (0.5)	83.7 (0.5)
0.05	0.95	-0.67 (0.425, -0.283)	85.7 (0.5)	85.5 (0.5)
0.1	0.75	-0.67 (0.237, -0.158)	81.5 (0.5)	81.4 (0.5)
0.1	0.95	-0.67 (0.300, -0.200)	82.0 (0.5)	81.9 (0.5)
0.0005	0.75	-0.50 (0.731, -0.366)	96.2 (0.3)	96.1 (0.3)
0.0005	0.95	-0.50 (0.926, -0.463)	99.9 (<0.1)	99.9 (<0.1)
0.05	0.75	-0.50 (0.335, -0.168)	83.9 (0.5)	83.7 (0.5)
0.05	0.95	-0.50 (0.425, -0.212)	86.2 (0.5)	86.2 (0.5)
0.1	0.75	-0.50 (0.237, -0.119)	81.7 (0.5)	81.8 (0.5)
0.1	0.95	-0.50 (0.300, -0.150)	83.2 (0.5)	82.9 (0.5)
0.0005	0.75	0.00 (0.731, 0.000)	96.1 (0.3)	95.9 (0.3)
0.0005	0.95	0.00 (0.926, 0.000)	100.0 -	100.0 -
0.05	0.75	0.00 (0.335, 0.000)	84.6 (0.5)	84.2 (0.5)
0.05	0.95	0.00 (0.425, 0.000)	87.1 (0.5)	86.8 (0.5)
0.1	0.75	0.00 (0.237, 0.000)	81.9 (0.5)	81.7 (0.5)
0.1	0.95	0.00 (0.300, 0.000)	83.5 (0.5)	83.2 (0.5)
0.0005	0.75	1.00 (0.731, 0.731)	96.5 (0.3)	96.5 (0.3)
0.0005	0.95	1.00 (0.926, 0.926)	99.9 (0.1)	99.9 (<0.1)
0.05	0.75	1.00 (0.335, 0.335)	84.1 (0.5)	84.0 (0.5)
0.05	0.95	1.00 (0.425, 0.425)	87.2 (0.5)	87.4 (0.5)
0.1	0.75	1.00 (0.237, 0.237)	82.6 (0.5)	82.6 (0.5)
0.1	0.95	1.00 (0.300, 0.300)	84.2 (0.5)	84.1 (0.5)
0.0005	0.75	2.00 (0.731, 1.462)	96.6 (0.3)	96.4 (0.3)
0.0005	0.95	2.00 (0.926, 1.852)	100.0 -	100.0 -
0.05	0.75	2.00 (0.335, 0.671)	84.9 (0.5)	85.4 (0.5)
0.05	0.95	2.00 (0.425, 0.850)	87.7 (0.5)	88.1 (0.5)
0.1	0.75	2.00 (0.237, 0.474)	83.4 (0.5)	83.9 (0.5)
0.1	0.95	2.00 (0.300, 0.601)	84.6 (0.5)	85.3 (0.5)

Table 8.30: Estimated Power with linear mixed effects model using a single covariate effect parameter (Adjusted) and separate covariate effects models (Separate effects). 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.05. (CEF=Covariate Effect Factor. RCE=Ratio of Covariate Effects. MCSE=Monte Carlo Standard Error.)

			Estimated Power % (MCSE)	
Covariate ICC	CEF	RCE (γ_w, γ_b)	Adjusted	Separate Effects
0.0005	0.75	-0.67 (0.712, -0.474)	94.1 (0.3)	94.0 (0.3)
0.0005	0.95	-0.67 (0.901, -0.601)	99.0 (0.1)	98.9 (0.1)
0.05	0.75	-0.67 (0.474, -0.316)	85.7 (0.5)	85.5 (0.5)
0.05	0.95	-0.67 (0.601, -0.401)	90.6 (0.4)	90.5 (0.4)
0.1	0.75	-0.67 (0.335, -0.224)	82.7 (0.5)	83.0 (0.5)
0.1	0.95	-0.67 (0.425, -0.283)	85.8 (0.5)	85.9 (0.5)
0.0005	0.75	-0.50 (0.712, -0.356)	93.9 (0.3)	93.6 (0.3)
0.0005	0.95	-0.50 (0.901, -0.451)	99.1 (0.1)	99.0 (0.1)
0.05	0.75	-0.50 (0.474, -0.237)	86.1 (0.5)	85.9 (0.5)
0.05	0.95	-0.50 (0.601, -0.300)	90.6 (0.4)	90.7 (0.4)
0.1	0.75	-0.50 (0.335, -0.168)	83.9 (0.5)	83.9 (0.5)
0.1	0.95	-0.50 (0.425, -0.212)	85.7 (0.5)	85.5 (0.5)
0.0005	0.75	0.00 (0.712, 0.000)	93.9 (0.3)	93.8 (0.3)
0.0005	0.95	0.00 (0.901, 0.000)	98.9 (0.1)	98.9 (0.1)
0.05	0.75	0.00 (0.474, 0.000)	87.1 (0.5)	87.1 (0.5)
0.05	0.95	0.00 (0.601, 0.000)	93.0 (0.4)	92.8 (0.4)
0.1	0.75	0.00 (0.335, 0.000)	84.9 (0.5)	84.7 (0.5)
0.1	0.95	0.00 (0.425, 0.000)	86.5 (0.5)	86.4 (0.5)
0.0005	0.75	1.00 (0.712, 0.712)	93.7 (0.3)	93.7 (0.3)
0.0005	0.95	1.00 (0.901, 0.901)	99.1 (0.1)	99.0 (0.1)
0.05	0.75	1.00 (0.474, 0.474)	87.7 (0.5)	87.6 (0.5)
0.05	0.95	1.00 (0.601, 0.601)	92.1 (0.4)	91.9 (0.4)
0.1	0.75	1.00 (0.335, 0.335)	84.9 (0.5)	84.8 (0.5)
0.1	0.95	1.00 (0.425, 0.425)	87.2 (0.5)	87.5 (0.5)
0.0005	0.75	2.00 (0.712, 1.423)	93.6 (0.3)	93.6 (0.3)
0.0005	0.95	2.00 (0.901, 1.803)	99.0 (0.1)	98.8 (0.2)
0.05	0.75	2.00 (0.474, 0.949)	88.5 (0.5)	89.4 (0.4)
0.05	0.95	2.00 (0.601, 1.202)	92.6 (0.4)	93.7 (0.3)
0.1	0.75	2.00 (0.335, 0.671)	84.8 (0.5)	86.5 (0.5)
0.1	0.95	2.00 (0.425, 0.850)	88.9 (0.4)	90.8 (0.4)

Table 8.31: Estimated Power with a linear mixed effects model using a single covariate effect parameter (Adjusted) or separate covariate effects models (Separate effects). 60 clusters per treatment arm, cluster size 5. Outcome ICC = 0.1. (CEF=Covariate Effect Factor. RCE=Ratio of Covariate Effects. MCSE=Monte Carlo Standard Error.)

8.6 Adjusting for cluster aggregated covariates

The cluster mean of an individual level covariate, rather than the covariate itself, could be adjusted for in a mixed effects model. In many of the simulations conducted, standard error was misestimated and estimated type I error was below the nominal level for this method of analysis. Some examples of estimated relative error of standard errors and estimate type I errors, from linear mixed effects model, are presented in Tables 8.32 and 8.33.

For a cluster size of five (Table 8.32, page 227), estimated type I error was as low as 0% for outcomes and covariates with ICCs of 0.0005, 0.5, and 0.1. Estimated relative error was as high as 186%, 161%, and 133% for outcomes and covariates with ICCs of 0.005, 0.05, and 0.1, respectively.

For a cluster size of 30 (Table 8.33, page 228), estimated type I error was as low as 0%, 0.1%, and 0.2% for outcomes and covariates with ICCs of 0.0005, 0.5, and 0.1, respectively. Estimated relative error was as high as 212%, 97%, and 51% for outcomes and covariates with ICCs of 0.005, 0.05, and 0.1, respectively.

Estimated relative errors increase, and type I errors decrease, as the effect of the covariate increases. There is greater misestimation of standard error when the outcome has a smaller ICCs. When the ICCs are 0.05 or 0.1, there is greater misestimation of standard error when the cluster size is smaller.

I undertook some simulations to further investigate this issue. Adjusting for a covariate that was generated at the cluster level did not lead to misestimation of standard errors. Adjusting for the true cluster means of an individual level covariate, that is that cluster level part of the generated covariate, plus an independent random error also did not lead to misestimation of standard errors. Some scenarios were also simulated using the R software (see Appendix E) to confirm that these results were not unique to Stata.

My investigations to date have not revealed the underlying cause for misestimation of the standard error of treatment effect, and estimated type I error below the nominal level, when adjusting for the cluster mean of a covariate. Further investigation of the cause of this problem is ongoing.

Outcome ICC	Covariate ICC	Covariate Effect Parameter	Estimated Relative Error of SE % (MCSE)	Estimated Type I Error % (MCSE)
0.0005	0.0005	0.00	3.3 (1.0)	4.3 (0.3)
0.0005	0.0005	0.25	3.5 (1.0)	4.4 (0.3)
0.0005	0.0005	0.50	13.4 (1.1)	3.1 (0.2)
0.0005	0.0005	0.75	43.2 (1.4)	0.5 (0.1)
0.0005	0.0005	0.95	186.3 (2.9)	0.0 -
0.05	0.05	0.00	0.0 (1.0)	5.0 (0.3)
0.05	0.05	0.25	1.8 (1.0)	4.8 (0.3)
0.05	0.05	0.50	4.3 (1.0)	4.6 (0.3)
0.05	0.05	0.75	28.8 (1.3)	1.4 (0.2)
0.05	0.05	0.95	161.3 (2.6)	0.0 -
0.1	0.1	0.00	0.3 (1.0)	5.1 (0.3)
0.1	0.1	0.25	-0.6 (1.0)	5.2 (0.3)
0.1	0.1	0.50	0.0 (1.0)	5.6 (0.3)
0.1	0.1	0.75	19.6 (1.2)	2.1 (0.2)
0.1	0.1	0.95	133.1 (2.3)	0.0 -

Table 8.32: Estimated relative error and type I error when using a linear mixed effects model, adjusted for a covariate aggregated by cluster mean. Cluster size of five.

Outcome ICC	Covariate ICC	Covariate Effect Parameter	Estimated Relative Error of SE % (MCSE)	Estimated Type I Error % (MCSE)
0.0005	0.0005	0.00	3.2 (1.0)	4.2 (0.3)
0.0005	0.0005	0.25	6.3 (1.1)	4.3 (0.3)
0.0005	0.0005	0.50	14.5 (1.1)	2.2 (0.2)
0.0005	0.0005	0.75	47.0 (1.5)	0.4 (0.1)
0.0005	0.0005	0.95	212.3 (3.1)	0.0 -
0.05	0.05	0.00	-1.9 (1.0)	5.4 (0.3)
0.05	0.05	0.25	-1.2 (1.0)	5.4 (0.3)
0.05	0.05	0.50	-0.9 (1.0)	5.7 (0.3)
0.05	0.05	0.75	2.6 (1.0)	5.3 (0.3)
0.05	0.05	0.95	96.9 (2.0)	0.1 -
0.1	0.1	0.00	1.2 (1.0)	5.2 (0.3)
0.1	0.1	0.25	0.4 (1.0)	6.4 (0.3)
0.1	0.1	0.50	0.4 (1.0)	5.7 (0.3)
0.1	0.1	0.75	-1.4 (1.0)	5.8 (0.3)
0.1	0.1	0.95	51.3 (1.5)	0.2 (0.1)

Table 8.33: Estimated relative error and type I error when using a linear mixed effects model, adjusted for a covariate aggregated by cluster mean. Cluster size of 30.

8.7 Conclusion

In this chapter I have presented results from simulation studies investigating the effects of covariate adjustment in the analysis of CRTs using mixed effects models.

Standard errors of treatment effect estimates were overestimated and estimated type I error was below the nominal level when adjusting for a covariate only when it is imbalanced at baseline. Standard errors were also overestimated and type I error was also below the nominal level when a covariate was used to stratify randomisation, but was not adjusted for in the analysis model.

In analyses using a linear mixed effects model, adjusting for a covariate with the same ICC as the outcome increased estimated power to a maximum of 100%. With a cluster size of five, adjusting for a covariate with a smaller ICC than the outcome generally increased estimated power more than adjusting for a cluster level covariate. However, with a cluster size of 30 and an outcome with an ICC of 0.05 or 0.1, adjusting for a cluster level covariate increased estimated power more than adjusting for a covariate with an ICC of 0.0005.

In analyses using a logistic mixed effects model with a cluster size of five, adjusting for a covariate with the same ICC as the outcome increased estimated power most notably. However, with a cluster size of 30 and an outcome with an ICC of 0.05 or 0.1 adjusting for a cluster level covariate often increased power as much. For an outcome with an ICC of 0.0005, adjusting for a covariate with an ICC of 0.05 or a cluster level covariate, did not significantly increase estimated power. With a cluster size of five and an outcome with an ICC of 0.05, adjusting for a covariate with a very small ICC increased estimated power more than adjusting for a cluster level covariate. But with a cluster size of 30, adjusting for a cluster level covariate increased estimated power more than adjusting for a covariate with an ICC of 0.0005.

In some circumstances, models using separate within-cluster and contextual covariate effect parameters gave greater power and precision than models with a single covariate effect paramter. However, under most scenarios models using separate or a single covariate effect parameter gave similar standard errors and power.

These results will be used in Chapter 10, along with an analytic investigation presented in the next chapter, to develop guidance for choosing covariates in the analysis of CRTs.

Chapter 9

Analytic investigation of power in adjusted linear mixed effects models

The results in Section 8.3 demonstrate that the ICC of the outcome, the ICC of the covariate, and cluster size all influence the potential increase in power when adjusting for a covariate. The estimates of power from simulations are consistent with power as calculated using a non-central F distribution 7.21. In this section, I use this method of estimating power to further investigate the effect of these parameters on the power of an adjusted analysis.

Recall the linear mixed effects model used for an adjusted analysis of a CRT:

$$Y_{ij} = \alpha + \beta X_j + \gamma Z_{ij} + u_j + e_{ij}$$

Y_{ij} is the outcome for individual i in cluster j and X_j indicates treatment arm ($X = 1$ for the experimental arm, $X = 0$ for the control arm). The u_j are independent cluster random effects with $u_j \sim N(0, \tau^2)$, and the e_{ij} are individual level residuals with $e_{ij} \sim N(0, \sigma^2)$. The power of the adjusted test of treatment effect can be calculated as

$$F_{tail}(1, J-3, \lambda, invF(1, J-3, 0.95))$$

where

$$\lambda = \frac{\beta^2 J}{4(\tau^2 + \sigma^2/m)} \ .$$

J is the number of clusters in the CRT. m is the number of individuals in each cluster. β is the treatment effect parameter. $invF(d_1, d_2, x)$ is the inverse cumulative F distribution function, with degrees of freedom d_1 and d_2. $F_{tail}(d_1, d_2, \lambda, x)$ is the

reverse cumulative (or upper tail) non-central F distribution function, with degrees of freedom d_1 and d_2, and non-centrality parameter λ.

Firstly, scale outcome variable Y and covariate Z to both have variance of one:

$$\begin{aligned}\gamma^2\tau_z^2 + \tau^2 + \gamma^2\sigma z^2 + \sigma^2 &= 1\\ \tau_z^2 + \sigma_z^2 &= 1\end{aligned}$$

Then we can relate the covariate effect and variance parameters in the model to the marginal ICCs of the outcome (ICC_Y) and covariate (ICC_Z) as follows:

$$\begin{aligned}\gamma^2\tau_z^2 + \tau^2 &= \text{ICC}_Y\\ \gamma^2\sigma_z^2 + \sigma^2 &= 1 - \text{ICC}_Y\\ \tau_z^2 &= \text{ICC}_Z\\ \sigma_z^2 &= 1 - \text{ICC}_Z\end{aligned}$$

The non-centrality parameter λ can then be written as

$$\lambda = \frac{1}{4}\beta^2 J\left(\text{ICC}_Y - \gamma^2\text{ICC}_Z + \frac{1 - \text{ICC}_Y - \gamma^2(1 - \text{ICC}_Z)}{m}\right)^{-1}.$$

A larger value of λ gives greater power for the adjusted analysis. Choosing covariates that will increase power the most depends on γ but also cluster size (m), the ICC of outcome (ICC_Y), and the ICC of covariate (ICC_Z). In particular, we want smallest value of

$$\frac{1}{4}\beta^2 J\lambda^{-1} = \text{ICC}_Y - \gamma^2\text{ICC}_Z + \frac{1 - \text{ICC}_Y - \gamma^2(1 - \text{ICC}_Z)}{m}.$$

Note that γ is covariate effect parameter when the total variances of the outcome and the covariate are both one. To consider the relationship between the outcome and covariate, it would be more useful to consider the linear correlation between them at the cluster and individual levels.

9.1 Comparing a cluster level, an individual level, and an unclustered covariate

I now consider three particular values for ICC_Z. If the covariate has the same marginal ICC as the outcome, then $\text{ICC}_Z = \text{ICC}_Y$. For a cluster level covariate, $\text{ICC}_Z = 1$. A covariate for which there is no clustering effect has $\text{ICC}_Z = 0$.

If $\text{ICC}_Z = \text{ICC}_Y$ then:

$$\begin{aligned}\frac{1}{4}\beta^2 J\lambda^{-1} &= \text{ICC}_Y(1 - \gamma^2) + \frac{(1 - \text{ICC}_Y)(1 - \gamma^2)}{m}\\ &= \text{ICC}_Y(1 - r^2) + \frac{(1 - \text{ICC}_Y)(1 - r^2)}{m}\end{aligned}$$

where r is the linear correlation between the outcome Y and covariate Z. Since $\text{ICC}_Z = \text{ICC}_Y$, this is the same correlation at cluster and individual level.

If $\text{ICC}_Z = 1$, that is Z is a cluster level covariate, then:

$$\begin{aligned}\frac{1}{4}\beta^2 J\lambda^{-1} &= \text{ICC}_Y - \gamma^2 + \frac{1-\text{ICC}_Y}{m} \\ &= \text{ICC}_Y(1-r_c^2) + \frac{(1-\text{ICC}_Y)}{m}\end{aligned}$$

Where r_c is the linear correlation between the covariate and true cluster means of the outcome variable.

If $\text{ICC}_Z = 0$ then:

$$\begin{aligned}\frac{1}{4}\beta^2 J\lambda^{-1} &= \text{ICC}_Y + \frac{(1-\text{ICC}_Y) - \gamma^2}{m} \\ &= \text{ICC}_Y + \frac{(1-\text{ICC}_Y)(1-r_i^2)}{m}\end{aligned}$$

Where r_i is the linear correlation between the covariate and individual level component of the outcome variable.

Firstly, compare adjusting for a cluster level covariate ($\text{ICC}_Z = 1$) and adjusting for individual level covariate with same ICC as outcome ($\text{ICC}_Z = \text{ICC}_Y$). Adjusting for either one covariate will give same power when:

$$\begin{aligned}&\text{ICC}_Y(1-r^2) + \frac{(1-\text{ICC}_Y)(1-r^2)}{m} = \text{ICC}_Y(1-r_c^2) + \frac{(1-\text{ICC}_Y)}{m} \\ \iff \quad & r^2\left(\text{ICC}_Y(m-1)+1\right) = r_c^2 \times m \times \text{ICC}_Y \\ \iff \quad & r^2\left(1 + \frac{1}{m}\left(\frac{1}{\text{ICC}_Y} - 1\right)\right) = r_c^2\end{aligned}$$

where r is the linear correlation between the outcome and the covariate with the same ICC, and r_c is the linear correlation between the cluster level covariate and true cluster means of the outcome variable. Where ICC_Y (=ICC_Z) is small, this equation can be approximated as:

$$r^2\left(1 + \frac{1}{\text{ICC}_Y \times m}\right) \approx r_c^2$$

For a large value of $(\text{ICC}_Y \times m)$, $\frac{1}{\text{ICC}_Y \times m} \approx 0$ so we have $r^2 \approx r_c^2$.

As $(\text{ICC}_Y \times m) \to 0$, $\frac{1}{\text{ICC}_Y \times m} \to \infty$ so

$$r^2\left(1 + \frac{1}{\text{ICC}_Y \times m}\right) > r_c^2$$

Therefore, for a large value of the product of cluster size and the ICC of the outcome, there is similar power adjusting for a covariate with the same ICC as the outcome or adjusting for a cluster level covariate. For a small value of the product of cluster

size and the ICC of the outcome, adjusting for the individual level covariate will give greater power.

Secondly, consider adjusting for either an individual level covariate with same ICC as outcome ($\text{ICC}_Z = \text{ICC}_Y$) or an individual level covariate with no clustering ($\text{ICC}_Z = 0$). Adjusting for either one covariate will give same power when:

$$\begin{aligned} & \text{ICC}_Y(1-r^2) + \frac{(1-\text{ICC}_Y)(1-r^2)}{m} = \text{ICC}_Y + \frac{(1-\text{ICC}_Y)(1-r_i^2)}{m} \\ \iff \quad & r^2\left(\text{ICC}_Y(m-1)+1\right) = r_i^2 \times (1-\text{ICC}_Y) \\ \iff \quad & r^2\left(1+m\left(\frac{\text{ICC}_Y}{1-\text{ICC}_Y}\right)\right) = r_i^2 \end{aligned}$$

Where ICC_Y (=ICC_Z) is small, this equation can be approximated as:

$$r^2\left(1+\text{ICC}_Y \times m\right) \approx r_i^2$$

Therefore, for a small value of the product of cluster size and the ICC of the outcome, there is similar power adjusting for a covariate with the same ICC as the outcome or an unclustered covariate. For a large value of the product of cluster size and the ICC of the outcome, adjusting for the covariate with the same ICC as the outcome will give greater power.

Finally, consider adjusting for either an individual level covariate with no clustering ($\text{ICC}_Z = 0$) or a cluster level covariate ($\text{ICC}_Z = 1$). Adjusting for either one covariate will give the same power if:

$$\begin{aligned} & \text{ICC}_Y(1-r_c^2) + \frac{(1-\text{ICC}_Y)}{m} = \text{ICC}_Y + \frac{(1-\text{ICC}_Y)(1-r_i^2)}{m} \\ \iff \quad & r_c^2 = \frac{1}{m}\left(\frac{1}{\text{ICC}_Y} - 1\right) r_i^2 \end{aligned}$$

Again, where ICC_Y is small, this is approximately:

$$r_c^2 \approx \frac{r_i^2}{\text{ICC}_Y \times m}$$

If the individual level covariate with no clustering and the cluster level covariate have the same correlation with the outcome variable at the individual and cluster level, respectively, ($r_i = r_c$), then adjusting for either covariate will give the same power if:

$$1 = (m+1) \times \text{ICC}_Y \approx m \times \text{ICC}_Y$$

If the product of ICC and cluster size is less than one, then adjusting for an unclustered covariate will give greater power than adjusting for a cluster level covariate with the same correlation with the outcome (albeit at the cluster level, rather than the individual level). If the product of ICC and cluster size is greater than one, then

adjusting for a cluster level covariate will give greater power than adjusting for an unclustered covariate that had same correlation with the outcome (at the individual level, rather than the cluster level).

Chapter 10

Choosing covariates in cluster randomised trials

In this chapter I address the second and third major aims of this thesis, namely:

2. To review published recommendations on choosing covariates, in the context of CRTs and identify which advice is applicable in the analysis of CRTs.
3. To develop further guidance for choosing covariates in the analysis of CRTs.

To do this I use the simulation results presented in Chapter 8 and the analytic work in Chapter 9. In Section 10.1 I consider existing guidance for choosing covariates in individually randomised trials, and discuss their relevance to CRTs in the light of simulation results. I discuss further recommendations specific to the analysis of CRTs in Section 10.2. In Section 10.3 I present a consolidated list of guidance for choosing covariates in the analysis of CRTs using linear and logistic mixed effects models.

10.1 Recommendations from individually randomised trial literature

A number of the recommendations for choosing covariates in individually randomised trials, which are described in Chapter 2, may be equally applicable to CRTs. I now consider several recommendations in turn, and review their validity for CRTs. I draw on the literature introduced in Chapter 3 and the results presented in Chapter 8.

10.1.1 Choosing covariates a priori

Previously published guidance [13, 14] that covariates to be adjusted for in a primary analysis of a trial should be chosen *a priori* is not specific to individually randomised trials. Raab & Butcher [23] argue that guidance for choosing covariates *a priori* also applies to planning primary analyses of CRTs. The following section discusses adjusting for a covariate due to an imbalance at baseline, but otherwise this thesis does not include any empirical evidence to support guidance to choose covariates *a priori*. Future investigation could possibly use simulations to investigate the validity of choosing covariates *post hoc*, for example using covariate selection algorithms or choosing covariates observed to be highly correlated with the outcome in the trial data, in the analysis of CRTs.

10.1.2 Adjusting for covariates significantly imbalanced between treatment arms

Adjusting for a covariate that is chosen *post hoc* because it is significantly imbalanced at baseline leads to overestimation of the standard errors of treatment effect estimates and type I error below the nominal level, as demonstrated in Section 8.1. This is evident for both individual and cluster level covariates. The same result is seen in the analysis of individually randomised trials [38, 39]. Therefore, guidance to not adjust for covariates due to an observed baseline imbalance is equally applicable to CRTs.

10.1.3 Covariates used in randomisation

The results in Section 8.2 show that failure to adjust for a covariate used to stratify randomisation leads to overestimation of the standard errors of treatment effect estimates and type I errors below the nominal level. This is demonstrated for cluster level covariates and individual level covariates aggregated to the cluster level. As in the analysis of individually randomised trials [13], to ensure a valid analysis models should be adjusted for covariates used to stratify randomisation.

10.1.4 Adjusting for a baseline measure of the outcome

The results in Section 8.3 and Chapter 9 show that the greatest increases in statistical power, and reductions in standard error for linear mixed effects models, are achieved when adjusting for a covariate with a high correlation with the outcome and the same ICC as the outcome. Baseline measures of an outcome variable are often highly

correlated with the outcome measured at follow-up. Further, we may reasonably expect that a baseline measure of the outcome variable will often have a similar ICC to the outcome measure at follow-up since it is the same variable being measured on the same sample. For example, in the OPERA and FIAT data sets used in Chapter 6 several outcome variables have similar ICCs at baseline and at follow-up. However, in some trials it may be expected that the ICC of the outcome will change notably between baseline and follow-up. In this case the guidance described later in Section 10.2 may be more useful. Broadly, though, the results presented in this thesis support guidance to adjust for a baseline measure of the outcome variable, where available, in the analysis of CRTs.

This concludes consideration of general guidance for choosing covariates in the analysis of individually randomised trials, and how they apply to the analysis of CRTs. In summary, guidance to choose covariates to be adjusted for *a priori*, adjust for covariates used to stratify randomisation, and adjust for a baseline measure of outcome where available is also applicable to the analysis of CRTs.

10.2 Recommendations for cluster randomised trials

I now consider how the results presented in Chapters 8 and 9 can be used provide further guidance for choosing covariates in the analysis of CRTs. These observations and results are specific to the analysis of CRTs.

10.2.1 Choosing covariates when using a linear mixed effects model

The results of simulations given in Section 8.3 and the work outlined in Chapter 9 are consistent, and show how the ICC of the outcome variable, the ICC of a covariate, and cluster size will influence the precision and power achieved in an adjusted analysis using a linear mixed effects model. Most relevant to the issue of choosing covariates in the analysis of CRTs is to identify under what conditions adjusting for particular covariates will most increase statistical power and precision of treatment effect estimates.

The work given in Chapter 9 identifies the product of the cluster size and the ICC of the outcome variable as a useful quantity to reveal what covariates may be most useful to increase precision and power. Alternatively, this product may be used to identify types of covariate where adjusting will have little or no effect in increasing precision or power. In general, if this product is close to one, then adjusting for either a covariate with an ICC of zero or a cluster level covariate (where the correlations

with the outcome at the individual and cluster level, respectively, are equal) will give similar power. For smaller values of this product, adjusting for a covariate with a similar or smaller ICC than the outcome will increase power and precision more than adjusting for a cluster level covariate or covariate with a larger ICC. For a larger value of this product, adjusting for a covariate with a similar or larger ICC than the outcome, such as cluster level covariates, will increase power and precision the most.

Although some authors have argued that cluster level covariates are most useful to improve precision of treatment effect estimates and power [23, 48, 56], this work shows that this is not necessarily true, as also demonstrated by Konstantopoulos [25]. This work goes further than Konstantopoulos [25] by showing how the ICC of the outcome and cluster size can be combined to be useful for choosing covariates in the analysis of CRTs.

10.2.2 Choosing covariates when using a logistic mixed effects model

In the analysis of a binary outcome using a logistic mixed effects model, adjusting for a cluster level covariate can increase precision and power. The adjusted treatment effect coincides with the unadjusted cluster specific treatment effect.

However, adjusting for an individual level covariate can increase power but reduce the precision of the estimate of treatment effect. The adjusted treatment effect in this case is conditional on the covariate, and is larger in magnitude than the unadjusted treatment effect. The greater the effect of the covariate on the outcome, the larger this difference is. This is equivalent to the known effect when adjusting for a covariate in the analysis of an individually randomised trial using a logistic regression model [5–7].

As with linear models, the ICC of the outcome and cluster size are useful to determine increases in power achievable by covariate with different ICCs. The expected value of the binary outcome, which is directly related to the variability of the outcome, has little effect on these considerations. When the product of cluster size and ICC of the outcome is large, similar increases in power are achieved when adjusting for a cluster level covariate or a covariate with the same ICC as the outcome. However, adjusting for a cluster level covariate also increases precision and does not alter the treatment effect that is being estimated. When the product of cluster size and ICC is small, a greater increase in power, but reduction in precision, can be achieved adjusting for a covariate with a similar or smaller ICC than the outcome.

10.2.3 Using separate cluster level and individual level covariate effects

The results given in Section 8.5 identify conditions where greater power and precision is achieved by using a model with separate within-cluster and contextual covariate effect parameters. Using a model with separate within-cluster and contextual covariate effect parameters can achieve greater power than a model with a single covariate effect parameter when there is a large difference between the two covariate effects, the overall effect of the covariate is large, the ICCs of the outcome and covariate are not too small, and the cluster size is not small. The precise conditions for these parameters to ensure greater power are not evident from this work, and would benefit from further investigation. However, these results suggest that for a cluster size of 30 the outcome and covariate should have ICCs of at least 0.05, and the between-cluster effect should be at least the same magnitude as the within-cluster effect (and so the contextual effect should be twice the magnitude of the within-cluster effect). For a cluster size of five, results suggest that there is minimal increase in power by using separate covariate effects for ICCs up to 0.1 and a between-cluster covariate effect up to twice the magnitude of the within-cluster covariate effect.

If there is *a priori* belief that the above conditions are fulfilled, then it may be beneficial to use a linear mixed effects model with separate within-cluster and contextual covariate effect parameters. A model with separate covariate effects may also be useful if the estimates of the covariate effects are of interest themselves, or if we require unbiased estimates of residual random effects [55].

10.2.4 Adjusting for cluster aggregated covariates

The results in Section 8.6 show that adjusting for an individual level covariate using only the cluster aggregated means may lead to misestimation of the standard error of the treatment effect estimate and type I error below the nominal level. This occurs particularly when cluster size and the ICCs of the outcome and covariate are small. Investigation so far has not found the underlying cause of this misestimation, and this would benefit from further investigation.

10.2.5 Choosing covariates in practice

The guidance presented in this thesis assumes some knowledge about the distributions of, and relationships between, outcome variables and covariates. However, when planning the analysis of CRTs, the true ICCs of the outcome and covariates, and

correlations between the outcome and covariates, are not known. The same issue is present in the planning of analyses of individually randomised trials. The true correlation between the outcome and covariates will not be known, but prognostic covariates can be identified using prior knowledge, and this is the recommended practice [13, 14]. Similarly, estimates of ICCs and correlations may be obtained from prior knowledge, for example from previous CRTs, other studies, or from a pilot trial. The CONSORT extension to CRTs [65] states that an estimated ICC (or coefficient of variation) should be reported for each primary outcome, providing a source of estimated ICCs for outcome variables. Reporting of ICCs for covariates in addition to this, for example in supplementary material or a database, may be valuable for those planning analyses of CRTs in the future.

10.3 Guidance for choosing covariates in the analysis of cluster randomised trials

1. Covariate adjustment can increase the power of an analysis of a CRT without a need to increase either the number of clusters or the number of individuals in the trial.

2. As for individually randomised trials, covariates to be adjusted for in the analysis of a CRT should be chosen *a priori*.

3. Neither individual level nor cluster level covariates should be adjusted for only because they are observed to be significantly imbalanced between treatment arms.

4. Covariates used to stratify randomisation should be used in a valid adjusted analysis.

5. Overall, adjusting for covariates that are highly correlated with the outcome variable will increase power most substantially.

6. Adjusting for a covariate with the same ICC as the outcome has the potential to increase power more than adjusting for a covariate with a different ICC.

7. A baseline measure of the outcome variable, where available, is likely to have a high correlation with the outcome measured at follow-up and a similar ICC. It may be particularly beneficial to adjust for a baseline measure of the outcome variable where available.

8. When the product of cluster size and the ICC of the outcome variable is close to one, power is similar when adjusting for a cluster level covariate or a covariate

with an ICC close to zero and similar correlation with the outcome.

9. When the product of cluster size and the ICC of the outcome variable is large, adjusting for a covariate with an ICC similar to or greater than the ICC of the outcome should be preferred in order to increase power most notably.

10. When the product of cluster size and the ICC of the outcome variable is smaller, adjusting for a covariate with an ICC similar to or smaller than the ICC of the outcome should be preferred in order to increase power most notably.

11. If the product of cluster size and outcome ICC is small, adjusting for cluster level covariates can only increase statistical power marginally.

12. When using a logistic mixed effects model, adjusting for an individual level covariate changes the treatment effect being estimated. However, adjusting for a cluster level covariate does not change the treatment effect being estimated.

13. When using a logistic mixed effects model, adjusting for an individual level covariate will typically increase the standard error of the treatment effect estimate. However, adjusting for a cluster level covariate will typically decrease the standard error of the treatment effect estimate.

14. If there is *a priori* belief that within-cluster and contextual covariate effect parameters differ substantially, the ICCs of the outcome and covariate are not too small, and the cluster size is not small, then using separate within-cluster and contextual covariate parameters may increase power further.

10.4 Conclusion

In this chapter I have directly addressed the question of how to choose covariates in the analysis of CRTs using linear and logistic mixed effects models. I used the simulation results in Chapter 8 and the analytic work presented in Chapter 9 to inform recommendations. Several of the guidelines for choosing covariates that were developed for individually randomised trials are also applicable to CRTs. I have produced guidance specific to the analysis of CRTs. In Section 10.3 I provide a summary of guidance, intended to be of practical use to those planning analyses of CRTs.

Chapter 11

Discussion and conclusion

Adjusting for covariates is common in primary analyses of CRTs, and can increase the statistical power of analysis without the need to increase the number of clusters or number of individuals in the trial. However, investigation of the effects of covariate adjustment and guidance for choosing covariates in the analysis of CRTs has been limited. This thesis addresses that gap in knowledge and guidance.

The three major aims of this project were:

1. To review existing knowledge and practice in the choice and handling of covariates in the analysis of CRTs.
2. To review published recommendations on choosing covariates, in the context of CRTs and identify which advice is applicable in the analysis of CRTs.
3. To develop further guidance for choosing covariates in the analysis of CRTs.

The first of these aims was addressed by a review of literature in Chapters 2 and 3, and by a review of CRTs which was described in Chapter 5. The review of literature for individually randomised trials identified extensive methodological research on the effects of covariate adjustment and consistent published guidance for choosing covariates. However, published research specific to the analysis of CRTs was found to be limited. The review of a large sample of CRT reports showed that adjusted analyses are common in CRTs, but there is very little reporting on how covariates are chosen. The review also showed poor adherence to some existing guidance, which is also the case in the results of previous reviews of individually randomised trials.

The second and third aims were addressed in Chapter 10, drawing on the results given in Chapters 8 and 9. Simulations were used to investigate the effects of covariate

adjustment in linear and logistic mixed effects models. The validity of adjusting for a covariate only when imbalanced at baseline, and adjusting for a covariate used to stratify randomisation, were also investigated. The use of separate within-cluster and contextual covariate effect parameters was also considered. An analytic investigation of power in adjusted linear mixed effects models extended the inquiry. These results were firstly used to review guidance for choosing covariates developed for individually randomised trials, in the context of the analysis of CRTs. The results were then used to review and extend guidance for choosing covariates that is specific to CRTs.

11.1 Contribution of this thesis to the field of research

There is extensive published research on the effects of covariate adjustment in the analysis of individually randomised trials, and there is clear guidance for choosing covariates in these trials. However, previously published work on the effects of covariate adjustment in CRTs is limited, focussing on linear mixed effects models and a limited number of other designs and analysis methods. In particular, there is minimal published work on the effects of covariate adjustment when using logistic mixed effects models.

This thesis presents work investigating the effects of covariate adjustment when using linear and logistic mixed effects in the analysis of CRTs with continuous or binary outcomes. The simulations for both linear and logistic mixed effects models are more extensive than those in previously published work.

This thesis also included a small analytic investigation of power in adjusted linear mixed effects models. This used the calculation for power described by Spybrook et al. [26] and Raudenbush et al. [48], but sought to produce some insight of practical use to researchers planning analyses of CRTs. I identified the product of cluster size and ICC of the outcome variable as a useful method for choosing covariates in the analysis of a CRT using mixed effects models.

A small number of authors have previously considered the use of separate within-cluster and contextual covariate effect parameters in linear mixed effects models [4, 54, 55]. The simulations presented in this thesis are more extensive than those in previously published work. In particular, I investigated a variety of values for ICCs, outcome variables and covariates with different ICCs from each other, and greater variety in the ratio of the within-cluster and contextual covariate effects than in previous investigations of power.

In Section 8.6, this thesis also highlighted a need for further investigation into the

effects of adjusting for an individual level covariate aggregated by cluster mean. In particular, I identified incorrect estimation of the standard error of treatment effect estimates when fitting such adjusted models.

The issue of choosing covariates in CRTs has rarely been addressed directly in previously published work, with little consideration for practical usefulness of results. I have shown that guidance for choosing covariates that was developed for individually randomised trials is also applicable to the analysis of CRTs. The guidance published by, for example, the European Medicines Agency [13] and the International Conference on Harmonization [15] should be equally heeded by researchers in the analysis of CRTs.

I have also used the simulation and analytic results presented in the thesis to produce guidance that is specific to CRTs, as given in Section 10.3. Although papers by, for example, Raab & Butcher [23], Raudenbush et al. [48], and Konstantopoulos [25] have partly addressed this question, consolidated guidance for choosing covariates in the analysis of CRTs has not been published to date.

11.2 Limitations of this work

In this thesis, I only considered the use of logistic and linear mixed effects models for the analysis of CRTs. However, other methods of analysis for CRTs such as generalised estimating equations are not considered. Additionally, analyses of count, categorical, and time-to-event outcomes have also not been considered. The work of this thesis is also limited to CRTs in which all clusters contain the same number of individuals, and there is an equal number of clusters in each treatment arm.

11.3 Further Work

There must be further investigation into the cause of incorrect estimation of standard error, and reduced type I error, under some conditions when adjusting for individual level covariates aggregated by cluster mean in mixed effects models. Appropriate methods for adjusting for covariates aggregated at the cluster level could then be developed.

CRTs with count, categorical, or time-to-event outcome variables are outside the scope of this work, but would benefit from similar investigation. In particular, it would be important to establish which of the guidance presented in this thesis is also applicable, or where further considerations must be taken into account. An investigation for

analysis of CRTs using generalised estimating equations would also be valuable.

This work could also be extended for application to CRTs with variable cluster size. If most clusters have the same size, and a minority of clusters have different sizes, it may be adequate to use the modal cluster size in place of cluster size for calculating the product of ICC and cluster size, to inform choice of covariates. Where cluster size is variable but in a narrow range, the utility of mean or median cluster size in the same way could be explored. If cluster size is widely variable then these findings may be of limited use.

An analytic investigation into the use of separate within-cluster and contextual covariate effect parameters would be valuable. In particular, it would be useful to identify precise conditions under which using separate within-cluster and contextual covariate effect parameters would give greater power than using a single covariate effect parameter. Additionally, an empirical investigation into the differences between within-cluster and contextual covariate effects that are found in practice would help to put that work into context.

11.4 Final remarks

This thesis has presented an investigation into the effects of covariate adjustment, in order to produce guidance for choosing covariates in the analysis of CRTs. I hope that the results and recommendations given in this thesis can be of some use to those planning analyses of CRTs. I also hope this work can institute further investigation of the use of covariates in the analysis of CRTs.

Appendix A

Glossary

Asymptotic bias

The difference between the treatment effect parameter in an adjusted model and the parameter in the unadjusted model (as per Gail et al. [6]).

Asymptotic relative efficiency (ARE)

Used to compare the efficiency of tests of no treatment effect. Suppose we have a parameter of interest θ and we wish to test $\theta = \theta_0$. And we have a statistic T that is a consistent estimator of $\mu_T(\theta)$, a monotone function (at least close to $\theta = \theta_0$) of θ. A definition of ARE called the Pitman efficiency of two such statistics T_1 and T_2 that are asymptotically normally distributed with means $\mu_{T_1}(\theta)$ and $\mu_{T_2}(\theta)$ and variances $\sigma^2_{T_1}$ and $\sigma^2_{T_2}$ is given by [28]:

$$ARE(T_1 \text{ to } T_2 \text{ at } \theta = \theta_0) = \left[\frac{\mu'_{T_1}(\theta_0)}{\mu'_{T_2}(\theta_0)}\right]^2 \left[\frac{\sigma^2_{T_2}(\theta_0)}{\sigma^2_{T_1}(\theta_0)}\right]$$

Asymptotic relative precision (ARP)

The *asymptotic relative precision (ARP)* of an estimator $\hat{\alpha}$ to an estimator $\hat{\beta}$ of a parameter is the ratio of the inverse of their variances:

$$ARP(\hat{\alpha} \text{ to } \hat{\beta}) = \frac{var(\hat{\alpha})^{-1}}{var(\hat{\beta})^{-1}} = \frac{var(\hat{\beta})}{var(\hat{\alpha})}$$

Bias

The expected difference between an estimator of a parameter and the true value of the parameter.

Binary (variable)
A random variable that can take one of two possible values, often zero or one.

Concave/convex (function)
A continuous function is convex if the value of the function at the midpoint of any interval is less than the mean of the values of the function at the ends of the interval. Equivalently, if the function has second derivative then it is convex if and only if that second derivative is greater than or equal to zero. A function f is concave if $-f$ is convex.

Confounder / confounding
A variable Z is a *confounder* of the relationship between exposure X and outcome Y if it is associated with exposure X and with outcome Y but is not on the causal pathway between X and Y.

Continuous (variable)
A random variable that can take any value between its minimum and maximum values.

Covariate
In this thesis, concerning randomised controlled trials, a *covariate* is a variable that is possibly related to the outcome variable and is independent of treatment arm allocation by randomisation.

Empirical standard error
The standard deviation of treatment effect estimates in a Monte Carlo sample.

Exchangeable
Random variables are *exchangeable* if any permutation of the variables has the same covariance structure.

Generalised linear model (GLM)
Generalised linear models (GLMs) are regression models that can be used to analyse a number of different types of outcome data including continuous and binary variables. A GLM has three components:

1. The probability distribution, from the exponential family, of the response variable Y.

2. A linear predictor, $\eta = \beta\mathbf{X}$ where $\mathbf{X}$ is a vector of predictor variables and β is a vector of parameters.

3. A link function g such that $E[Y] = \mu = g^{-1}(\eta)$.

Common link functions are linear, log, logit, probit, complementary log-log, and generalised logistic ($g(\mu) = \log(\mu^\theta/(1-\mu^\theta))$) link functions.

Independence
Two random variables are independent if and only if their joint cumulative distribution function is the product of their cumulative distribution functions.

Individually randomised trial
A randomised trial in which individuals are randomised to treatment arms. The randomisation unit (individuals) is the same as the unit of analysis (individuals)

Jensen's inequality
If f is a convex function and X is a random variable, then

$$E[f(X)] \geq g(E[X]) \ .$$

If f is a concave function, then

$$E[f(X)] \leq g(E[X]) \ .$$

Likelihood
The probability of an observed outcome given values for parameters.

Likelihood ratio test
A statistical test using the ratio of likelihoods, used to compare the fit of two models.

Linear predictor
A linear combination of parameters and predictor variables.

Link function

The function that relates the expected value of the outcome to the linear predictor in a generalised linear model.

Maximum likelihood estimator (MLE)

The value of the paramter than maximises the likelihood.

Minkowski's inequality

If $p > 1$ then:

$$\left(\sum_{k=1}^{n}(a_k + b_k)^p\right)^{1/p} \leq \left(\sum_{k=1}^{n} a_k^p\right)^{1/p} + \left(\sum_{k=1}^{n} b_k^p\right)^{1/p}$$

If $p < 1$, $p \neq 0$, then:

$$\left(\sum_{k=1}^{n}(a_k + b_k)^p\right)^{1/p} \geq \left(\sum_{k=1}^{n} a_k^p\right)^{1/p} + \left(\sum_{k=1}^{n} b_k^p\right)^{1/p}$$

Monte Carlo standard error

The estimated standard error of parameter estimates produced by Monte Carlo simulation.

Parameter

An, often unknown, value that describes a characteristic of the population or relationship between variables.

Power

The probability that a statistical test correct rejects the null hypothesis.

Precision

The inverse of the variance of an estimator.

Relative error of model-based standard error

A comparison of the mean model-based standard error and the empirical standard error:

$$\frac{\bar{s}}{\text{Empirical Standard Error}} - 1$$

where $\bar{s}$ is a mean of the model-based standard errors in a Monte Carlo sample.

Taylor series expansion

The Taylor series expansion of a real function $f(x)$ at $x = a$ is:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \dots$$

Variance

The expected value of the square of the difference between a random variable and the expected value of the random variable. Equal to the square of the standard deviation.

Wald test

A significance test of a value of a parameter. The Wald statistic for an estimator $\hat{\theta}$ of the parameter θ_0,

$$\frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$$

is compared to the standard normal distribution, or a t distribution.

Appendix B

Notation

General

Greek letters, e.g.ξ, η	Parameters
Hatted letters, e.g. $\hat{\beta}$	Estimators of parameters
Capital letters, e.g. X, Y	Random variables
Bold letters, e.g. $\mathbf{v}$, $\mathbf{A}$	Vectors and matrices
$pr(.)$, $P(.)$	Probability
$\bar{X}$	Mean of X
$E[X]$	Expected value of X
$Var(X)$	Variance of X
σ	Standard deviation parameter
σ^2	Variance parameter
Ω	Covariance matrix
η	Linear predictor for GLM
$g(\eta)$	Link function for GLM
$g^{-1}(\eta) = h(\eta)$	Inverse link function for GLM

Individually randomised parallel group trials

Y	Outcome variable
X	Binary treatment arm variable
Z	Covariate
α	Intercept parameter
β	Treatment effect parameter, adjusted
β^*	Treatment effect parameter, unadjusted
γ	Covariate effect parameter
$\psi = \gamma Z$	Covariate effect

Cluster randomised trials

Y	Outcome variable
X	Binary treatment arm variable
Z	Covariate
α	Intercept parameter
β	Treatment effect parameter, adjusted
β^*	Treatment effect parameter, unadjusted
γ	Covariate effect parameter
γ_W	Within-cluster covariate effect parameter
γ_C	Contextual covariate effect parameter
u_j	Cluster random effects
e_{ij}	Individual residuals
τ^2	Cluster level variance
σ^2	Individual level variance
$\tau^2_{Y\|Z}$	Residual cluster level variance of Y
$\sigma^2_{Y\|Z}$	Residual individual level variance of Y
a_j	Cluster level random effect of covariate
b_{ij}	Individual level random effect of covariate
τ^2_z	Cluster level variance of Z
σ^2_z	Individual level variance of Z
ICC_Y, ICC_Z	Intra-cluster correlation coefficient of Y, Z
J, indexed by j	Number of clusters
m, indexed by i	Cluster size

Appendix C

Details of literature on effects of covariate adjustment in individually randomised trials

C.1 Approach of Gail et al. using moments

In this section I present details of the work of Gail et al. [6].

Gail et al. [6] first define adjusted and unadjusted models for the expectation of the outcome, as follows: Let Y be an outcome variable, X a binary variable identifying treatment arm (taking value 1 with probability p) and Z a covariate (which may be a vector of covariates, although most results are derived as if it is a scalar). By randomisation, X and Z are independent. Assume that $E(Z) = 0$ and $E(Z^2) = \sigma^2$ (or where Z is a vector use Ω for the covariance matrix).

Firstly, consider the model that includes a covariate effect. Assume the conditional expectation of Y given X and Z satisfies

$$E(Y|X, Z) = h(\alpha + \beta X + \gamma Z) = h(\eta) = g^{-1}(\eta)$$

where $h(.)$ is a known function. So β is the treatment effect parameter adjusted for the covariate Z. Given this model, we have the following moments:

$$\begin{aligned} E(Y) &= E\{h(\alpha + \beta X + \gamma Z)\} \\ E(XY) &= E\{Xh(\alpha + \beta X + \gamma Z)\} \\ E(ZY) &= E\{Zh(\alpha + \beta X + \gamma Z)\} \end{aligned} \tag{C.1}$$

If these are N observations (i.e. patients in the trial) then the method of moment

estimators for the parameters $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ are the solutions to:

$$\begin{aligned} N^{-1}\sum y_i &= N^{-1}\sum h(\hat{\alpha} + \hat{\beta}x_i + \hat{\gamma}z_i) \\ N^{-1}\sum x_i y_i &= N^{-1}\sum x_i h(\hat{\alpha} + \hat{\beta}x_i + \hat{\gamma}z_i) \\ N^{-1}\sum z_i y_i &= N^{-1}\sum z_i h(\hat{\alpha} + \hat{\beta}x_i + \hat{\gamma}z_i) \end{aligned} \tag{C.2}$$

Now consider the unadjusted model, that does not include a covariate effect. So it is assumed instead that the conditional expectation of Y given X and Z satisfies

$$E(Y|X,Z) = E(Y|X) = h(\alpha^* + \beta^* X) = h(\eta^*) = g^{-1}(\eta^*)$$

where $h(.)$ is a known function. So β^* is the unadjusted treatment effect parameter. Given this model, we have the moments:

$$\begin{aligned} E(Y) &= E\{h(\alpha^* + \beta^* X)\} \\ E(XY) &= E\{Xh(\alpha^* + \beta^* X)\} \end{aligned} \tag{C.3}$$

Then the method of moments estimators of the unadjusted parameters $(\hat{\alpha}^*, \hat{\beta}^*)$ would be the solutions to:

$$\begin{aligned} N^{-1}\sum y_i &= N^{-1}\sum h(\hat{\alpha}^* + \hat{\beta}^* x_i) \\ N^{-1}\sum x_i y_i &= N^{-1}\sum x_i h(\hat{\alpha}^* + \hat{\beta}^* x_i) \end{aligned} \tag{C.4}$$

Now we can define the *asymptotic bias* of the unadjusted treatment effect by the difference between the unadjusted and adjusted treatment effect parameters: $\delta = \beta^* - \beta$.

The following Theorem C.1 shows from the methods of moments equations that there is no asymptotic bias if at least one of three conditions hold: there is no treatment effect, there is no covariate effect or the covariate does not vary.

Theorem C.1 (Corollary 1 of Gail et al. [6]).
Given the models and definitions defined above. The adjusted and unadjusted parameters of treatment effect coincide ($\delta = 0$) if $\beta = 0$ or $\gamma = 0$ or if Z does not vary ($\sigma^2 = 0$).

Proof. If $\gamma = 0$ then equations(C.3 are immediately identical to equations C.1. If $\beta = 0$ then $\xi_1 = \xi_0$ so $\beta^* = 0 = \beta$. If Z does not vary then equations C.3 and C.1 are identical with $\beta^* = \beta$ and $\alpha^* = \alpha + \gamma Z$. □

Define the following expected values of the outcome Y conditional on treatment arm:

$$\begin{aligned} & \xi_1 = E(Y|X=1) = E_Z(h(\alpha+\beta+\gamma Z)) \\ \text{and} \qquad & \xi_0 = E(Y|X=0) = E_Z(h(\alpha+\gamma Z)) \end{aligned}$$

We can then express the unadjusted parameters (α^* and β^*) in terms of the adjusted parameters (α, β and γ) by equating equations C.3 with the first two equations of C.1:

$$\begin{aligned} \alpha^* &= h^{-1}(\xi_0) \\ \beta^* &= h^{-1}(\xi_1) - h^{-1}(\xi_0) \end{aligned} \tag{C.5}$$

We require $h(\eta)$ to have the unique inverse h^{-1} that is well defined at ξ_1 and ξ_0.

By a second order Taylor expansion of the expression of the unadjusted treatment effect parameter β^* given in equations C.5, Gail et al. [6] give an approximate expression for the asymptotic bias, for a covariate effect close to zero. This is shown in Theorem C.2. From this expression, it is shown that there is no asymptotic bias if the function of the linear predictor giving the expected value of the outcome (the function $h(.)$ is linear or exponential (Theorem C.3). Finally, in Theorem C.4 it is shown that using a GLM and maximum likelihood estimators gives the same results.

Theorem C.2 (Theorem 2 of Gail et al. [6]).
Given the conditions of Theorem C.1 and also assume $h'(.)$ and $h''(.)$ exist, h^{-1} is not singular at $h(\alpha+\beta)$ and $h(\alpha)$, and ξ_1 and ξ_0 are well approximated, for γ close to zero, by the second-order Taylor series:

$$\begin{aligned} \xi_1 &= E_Z(h(\alpha+\beta+\gamma Z)) = h(\alpha+\beta) + \tfrac{1}{2}\gamma'\Omega\gamma(\alpha+\beta) \\ \xi_0 &= E_Z(h(\alpha+\gamma Z)) = h(\alpha) + \tfrac{1}{2}\gamma'\Omega\gamma(\alpha) \end{aligned}$$

Then, for γ close to zero, the asymptotic bias of the unadjusted treatment effect is

$$\delta = \tfrac{1}{2}\gamma'\Omega\gamma \left\{ \frac{h''(\alpha+\beta)}{h'(\alpha+\beta)} - \frac{h''(\alpha)}{h'(\alpha)} \right\} .$$

Proof. This is obtained from the second-order Taylor expansion of β^* in equation C.5. □

Theorem C.3 (Theorem 5 of Gail et al. [6]).
Assuming the same conditions as Theorems C.1 and C.2 (and β, γ and σ^2 are non-zero), the adjusted and unadjusted parameters of treatment effect coincide ($\delta = 0$) if and only if $h(\eta) = a\eta + b$ or $h(\eta) = ce^{a\eta+b} + d$.

Proof.

$$\begin{aligned} & \delta = 0 \\ \Leftrightarrow \qquad & \frac{h''(\alpha+\beta)}{h'(\alpha+\beta)} - \frac{h''(\alpha)}{h'(\alpha)} = 0 \\ \Leftrightarrow \qquad & \frac{h''(\alpha+\beta)}{h'(\alpha+\beta)} = \frac{h''(\alpha)}{h'(\alpha)} \end{aligned}$$

for arbitrary values of α and β. So $\delta = 0$ if and only if the function $\frac{h''(.)}{h'(.)}$ is constant.

$\frac{h''(.)}{h'(.)} = 0$ if and only if $h(\eta) = a\eta + b$ by integrating or differentiating twice. $\frac{h''(.)}{h'(.)} = a \neq 0$ if and only if $h(\eta) = ce^{a\eta+b} + d$ as the general solution to the differential equation $h''(\eta) - ah'(\eta) = 0$. □

Theorem C.4 (Theorem 6 of Gail et al. [6]).
Consider the GLM defined by

$$E(Y|X, Z) = g^{-1}(\eta) = h(\eta) \ ,$$
$$\eta = \alpha + \beta X + \gamma Z$$

and Y follows a distribution in the exponential family, so that β is the treatment effect adjusted for the covariate Z. Also consider an alternative GLM with linear predictor

$$\eta^* = \alpha^* + \beta^* X$$

so that β^* is the unadjusted treatment effect. Assume $\theta'(\alpha^* + \beta^*)$ and $\theta'(\alpha^*)$ do not vanish, where θ is the canonical parameter of the distribution of Y (regarded as a function of the linear predictor). Then the solutions for α^* and β^* are given by the equations in C.5, and the results of Theorems C.1, C.2 and C.3 also apply to these parameters.

Proof. Recall the log likelihood for a distribution in the exponential family is

$$l(\theta; y) = \frac{y\theta - b(\theta)}{a(\psi)} + c(y, \psi)$$

Assume $a(\psi)$ is known, as maximum likelihood estimates of $a(\psi)$ are asymptotically independent of the estimates of α, β and γ. Also recall

$$E(Y|X, Z) = g^{-1}(\eta) = h(\eta) = b'(\theta)$$

and regard the canonical parameter θ as a function of the linear predictor η.

Under the adjusted GLM, the maximum likelihood equations for N observations are

$$\begin{aligned} \sum_{i=1}^{N} \theta'(\eta_i) \left(y_i - h(\eta_i)\right) &= 0 \\ \sum_{i=1}^{N} \theta'(\eta_i) x_i \left(y_i - h(\eta_i)\right) &= 0 \\ \sum_{i=1}^{N} \theta'(\eta_i) z_i \left(y_i - h(\eta_i)\right) &= 0 \end{aligned} \tag{C.6}$$

Multiplied by N^{-1}, equations C.6 converge to

$$E\left[\theta'(\eta)(Y - h(\eta))\right] = 0 \ ,$$
$$E\left[\theta'(\eta)X(Y - h(\eta))\right] = 0 \ ,$$
$$E\left[\theta'(\eta)Z(Y - h(\eta))\right] = 0 \ .$$

Similarly, under the unadjusted GLM, the maximum likelihood equations converge to

$$E\left[\theta'(\eta^*)(Y - h(\eta^*))\right] = 0 \ , \qquad E\left[\theta'(\eta^*)X(Y - h(\eta^*))\right] = 0 \ .$$

For p as the probability of being assigned to the active treatment arm, these equations are equivalent to

$$p\theta'(\alpha^* + \beta^*)\xi_1 + (1-p)\theta'(\alpha^*)\xi_0 = p\theta'(\alpha^* + \beta^*)h(\alpha^* + \beta^*) + (1-p)\theta'(\alpha^*)h(\alpha^*)$$
$$p\theta'(\alpha^* + \beta^*)\xi_1 = p\theta'(\alpha^* + \beta^*)h(\alpha^* + \beta^*)$$

If $\theta'(\alpha^* + \beta^*)$ and $\theta'(\alpha^*)$ do not vanish, then the solutions to these equations are exactly as given in equations C.5. So, the results of Theorems C.1, C.2 and C.3 also apply to these parameters. □

C.2 Geometric approach of Neuhaus & Jewell

In this section I present details of the work of Neuhaus & Jewell [7].

Let Y be an outcome variable, X a binary variable identifying treatment arm (taking value 1 with probability $p = \frac{1}{2}$) and Z a covariate (here, we consider Z as one-dimensional but the methods generalise to Z as a multi-dimensional vector). By randomisation, X and Z are independent. Assume that $E(Z) = \mu_Z$ and $var(Z) = \sigma_Z^2$. Consider the GLM that includes the covariate, defined by

$$E(Y|X,Z) = g^{-1}(\eta) = h(\eta) \ ,$$
$$\eta = \alpha + \beta X + \gamma Z$$

and Y follows a distribution in the exponential family. So β is the adjusted treatment effect parameter. Assume the link function g is strictly monotone increasing and differentiable. For convenience, define the *covariate effect* $\psi = \gamma Z$ so $E(\psi) = \gamma\mu_Z$ and $var(\psi) = \gamma^2\sigma_Z^2$.

The 'geometric' method of Neuhaus & Jewell [7] is based on the relationship between: (1) the expected value of the outcome in the control treatment arm, denoted by μ_0, and (2) the absolute difference between expected values of the outcome in each treatment arm, denoted by $\Delta = \mu_1 - \mu_0$. So, formally, we have

$$\mu_k = E(Y|X = k, \psi) = g^{-1}\{\alpha + \beta k + \psi\} \qquad \text{for } k = 0, 1$$
$$\text{and} \qquad \Delta = \mu_1 - \mu_0 \ .$$

As per Neuhaus & Jewell [7], we can express $\Delta(\psi)$ in terms of $\mu_0(\psi)$ and β:

$$\begin{aligned} \Delta(\psi) = \mu_1 - \mu_0 &= g^{-1}\{\alpha + \beta + \psi\} - \mu_0(\psi) \\ &= g^{-1}\{g(\mu_0(\psi)) + \beta\} - \mu_0(\psi) \end{aligned} \tag{C.7}$$

Then the adjusted treatment effect parameter is $\beta = g(\mu_1) - g(\mu_0)$.

Under this adjusted GLM, the expected value of the outcome in each treatment arm, μ_0 and μ_1, varies for different values of the covariate effect, ψ. For example, suppose we use a logistic regression model with a treatment effect parameter of 1.3 and intercept parameter of -1. Figure C.1a shows the expected value of outcome in each treatment arm (μ_1 and μ_0) plotted against the covariate effect. Figure C.1b shows the same plot but using an identity link function (and same parameter values). Finally, Figure C.1c shows the same plot but using a log link function. In each of the plots, the horizontal difference between the two curves is constant, as the treatment effect parameter is constant and acts on the same scale as the covariate effect (the scale of the linear predictor). The vertical difference between the curves is the absolute difference between expected values of the outcome in each treatment arm, Δ. We can see that for logistic regression, this difference varies for different values of the covariate effect. For the identity link function, the difference is constant for all values of the covariate effect. And for a log link function the difference is proportional to the expected value of the outcome in the control arm.

We can instead plot this absolute difference Δ against the expected value of the outcome in the control treatment arm μ_0, as described by Neuhaus & Jewell [7]. The shape of this curve depends on the link function used in the GLM and the value of the treatment effect parameter. Link function g is assumed increasing, so g^{-1} is also an increasing function and the (μ_0, Δ) curves are nested as β increases ($\frac{\partial \Delta}{\partial \beta} > 0$ for $\beta > 0$). Figure C.2a shows such curves for the logit link function and treatment effect parameter values of 1.5, 0.5 and 0.05. Figures C.2b and C.2c show such curves for the identity and log link function (and the same treatment effect parameter values).

Now consider omitting the covariate Z, so we use the same GLM except with linear predictor

$$\eta^* = \alpha^* + \beta^* X$$

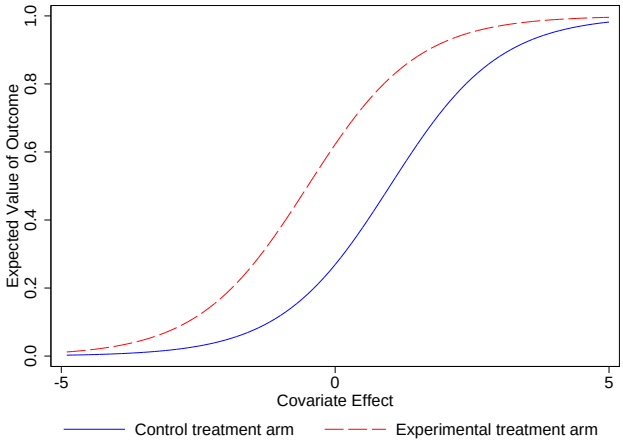
so that β^* is the unadjusted treatment effect parameter. Averaging with respect to ψ (i.e. averaging over the distribution of the omitted covariate Z) gives

$$E_\psi[Y|X=1] - E_\psi[Y|X=0] = E_\psi[\mu_1] - E_\psi[\mu_0]$$

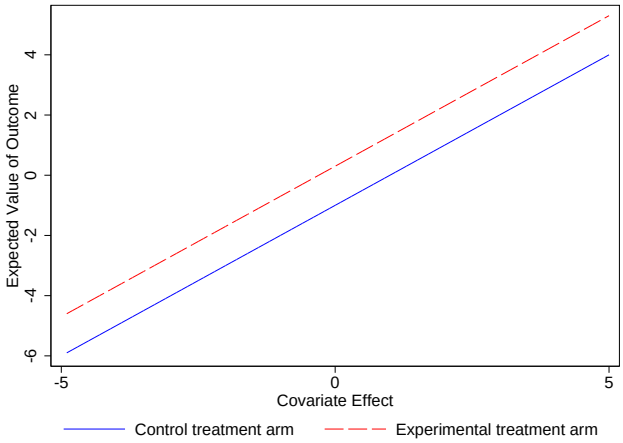
and so $\beta^* = g(\mu_1^*) - g(\mu_0^*)$ with $\mu_k^* = E_\psi[\mu_k(\psi)]$. We can average over the marginal density of Z since X and Z are independent by design.

Averaging equation C.7 with respect to ψ gives $\Delta^* = E_\psi[\Delta] = \mu_1^* - \mu_0^*$. The point (μ_0^*, Δ^*) in the (μ_0, Δ) plane is found, then, by averaging the (μ_0, Δ) curve with respect to the distribution of ψ. Here, μ_0^* is the expected value of the outcome in the

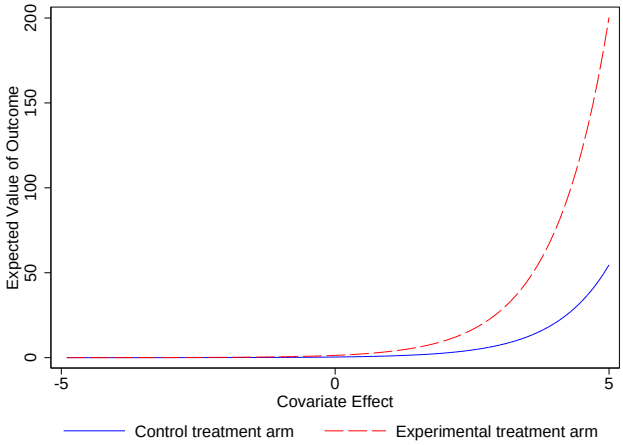
control treatment arm under the unadjusted model, and Δ^* is the absolute difference between expected values of the outcome in each treatment arm under the unadjusted model. For example, assuming again a treatment effect parameter of 1.3 and intercept parameter of -1, and also assume that the covariate Z is binary (taking value 1 with probability 3/4) with effect parameter of 1.6. Figure C.3 shows the (μ_0, Δ) curves and (μ_0^*, Δ^*) points in this case when using either a logistic, identity or log link function.



(a) With logit link function

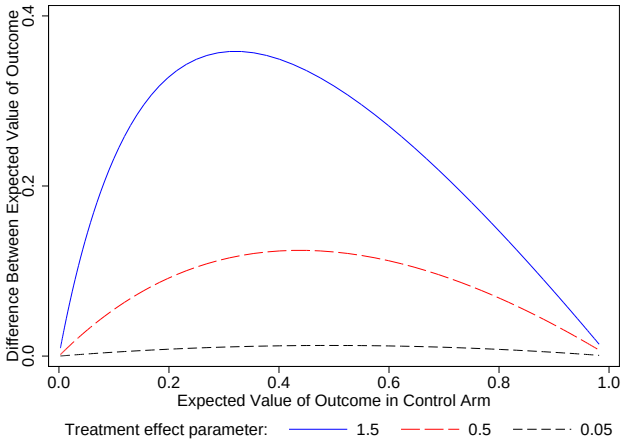


(b) With identity link function

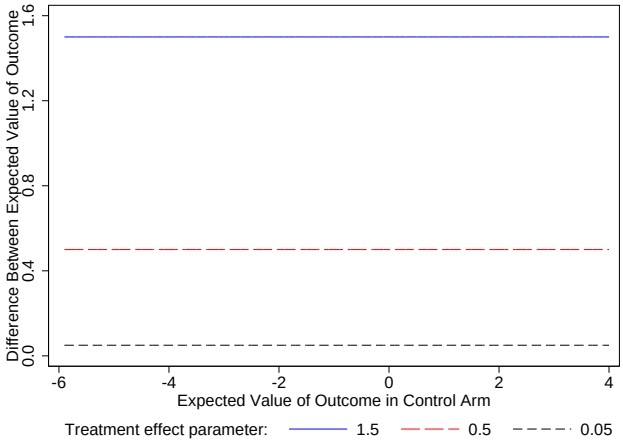


(c) With log link function

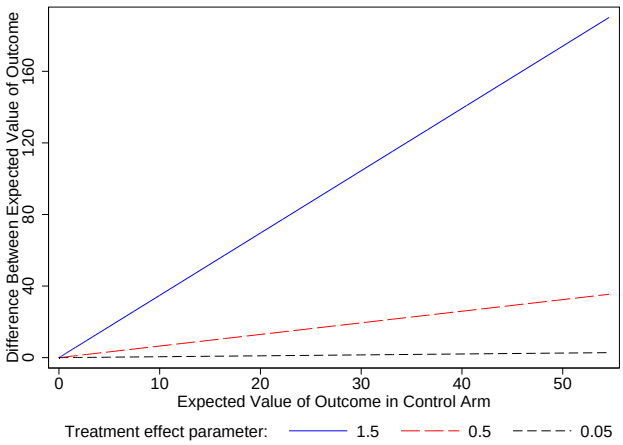
Figure C.1: Expected value of outcome in each treatment arm (μ_1 and μ_0) plotted over covariate effect ψ (with treatment effect parameter of 1.3 and intercept parameter of -1).



(a) With logit link function

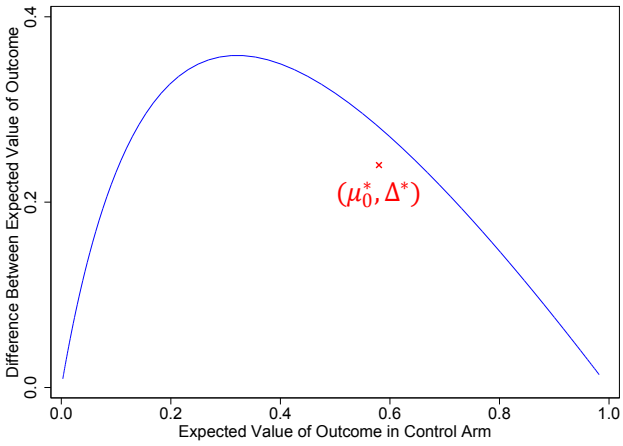


(b) With identity link function

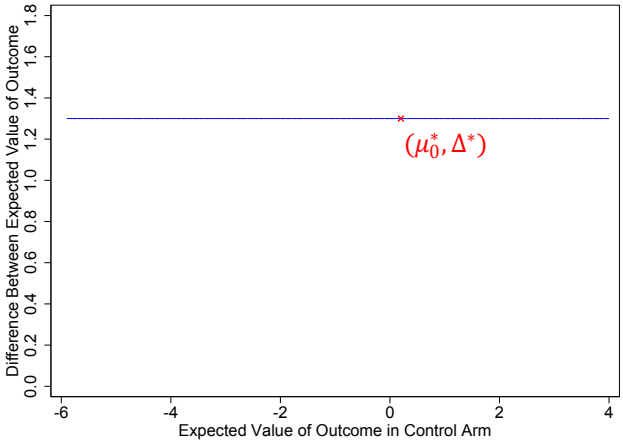


(c) With log link function

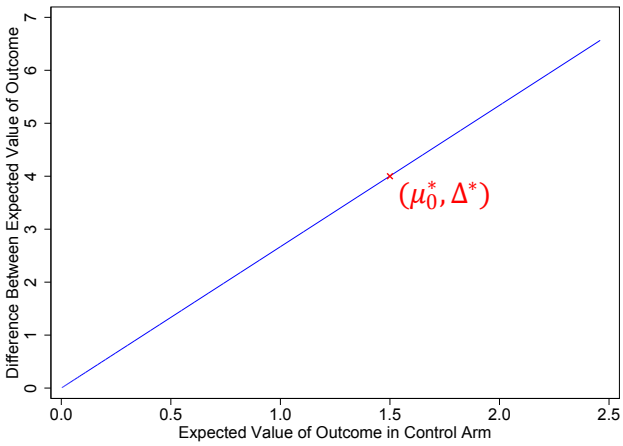
Figure C.2: Difference between expected values of the outcome in each treatment arm, Δ, plotted against the expected value of the outcome in the control treatment arm, μ_0. For treatment effect parameters, $\beta = 1.5, 0.5, 0.05$.



(a) With logit link function



(b) With identity link function



(c) With log link function

Figure C.3: Difference between expected values of the outcome in each treatment arm, Δ, plotted against the expected value of the outcome in the control treatment arm, μ_0. And the point (μ_0^*, Δ^*), which is these expected values averaged over the distribution of the binary covariate Z (taking value 1 with probability 3/4) with effect parameter of 1.6. For treatment effect parameter, $\beta = 1.3$.

The following two lemmas, theorem and corollary find how the direction of the asymptotic bias depends on properties of the link function used in the GLM. Lemma C.5 shows that the curve that passes through the point (Δ^*, μ_0^*) is given by the unadjusted treatment effect parameter β^*. Thus, the position of the averaged absolute treatment difference Δ^*, relative to the curve given by the adjusted model will determine the direction of the asymptotic bias. Lemma C.6 shows by Jensen's inequality that the concavity of the (μ_0, Δ) curve determines the position of the point (μ_0^*, Δ^*) with respect to the curve, and thus the direction of the asymptotic bias. Theorem C.7 then relates the direction of asymptotic bias directly to the concavity of a function defined via link function. Corollary C.8 gives the direction of asymptotic bias for linear, log and logit link functions.

Lemma C.5.
The curve (μ_0, Δ) (with $\beta = \beta_0$) passes through the point (μ_0^*, Δ^*) if and only if $\beta_0 = \beta^*$.

Proof. By the monotonicity of g and the continuity of g^{-1}, there exists a curve (μ_0, Δ) of the form of equation C.7 which passes through (μ_0^*, Δ^*) for some $\beta = \beta_0$. To find β_0, substitute $\Delta = \Delta^*$ and $\mu_0 = \mu_0^*$ into equation C.7:

$$\begin{aligned} & \Delta^* = g^{-1}\{g(\mu_0^*) + \beta_0\} - \mu_0^* \\ \Leftrightarrow \quad & \mu_1^* - \mu_0^* = g^{-1}\{g(\mu_1^*) - \beta^* + \beta_0\} - \mu_0^* \\ \Leftrightarrow \quad & g(\mu_1^*) = g(\mu_1^*) - \beta^* + \beta_0 \\ \Leftrightarrow \quad & \beta_0 = \beta^* \end{aligned}$$

So the curve (μ_0, Δ) (with $\beta = \beta_0$) passes through the point (μ_0^*, Δ^*) if and only if $\beta_0 = \beta^*$. □

Lemma C.6.
If the curve (μ_0, Δ) is concave, convex or linear then $\beta^* \leq \beta$, $\beta^* \geq \beta$ or $\beta^* = \beta$ respectively. If the curve (μ_0, Δ) is strictly concave or convex then $\beta^* = \beta$ if and only if $\gamma = 0$ or $var(Z) = 0$.

Proof. By Lemma C.5, the position of the point (μ_0^*, Δ^*) compared to the curve (μ_0, Δ) will tell us the relationship between β and β^*.

By Jensen's inequality, if $\Delta(\mu_0)$ is a concave function, then

$$\Delta[E_\psi(\mu_0)] \geq E_\psi[\Delta]$$

and so

$$\Delta[\mu_0^*] \geq \Delta^* \ .$$

And so the point (μ_0^*, Δ^*) lies below the curve (μ_0, Δ), and since Δ is an increasing function of β we have $\beta \geq \beta^*$.

Similarly if $\Delta(\mu_0)$ is a convex function then $\beta \leq \beta^*$, and if $\Delta(\mu_0)$ is linear then $\beta = \beta^*$

If $\Delta(\mu_0)$ is strictly concave or convex, then $\beta = \beta^*$ if and only if μ_0 does not vary, that is if and only if $\gamma = 0$ or $var(Z) = 0$. □

Theorem C.7.
If $\beta = 0$ then $\beta^* = \beta = 0$, otherwise if $F(t) = -1/g'(t)$ is strictly convex, concave or linear then $|\beta^*| < |\beta|$, $|\beta^*| > |\beta|$ or $\beta^* = \beta$ respectively.

Proof. By evaluating derivatives of Δ

$$\frac{\partial^2 \Delta}{\partial \mu_0^2} = \frac{g''(\mu_0)(g'[g^{-1}\{g(\mu_0)+\beta\}])^2 - \{g'(\mu_0)\}^2 g''[g^{-1}\{g(\mu_0)+\beta\}]}{(g'[g^{-1}\{g(\mu_0)+\beta\}])^3} .$$

If $\beta = 0$ then this second derivative is zero and the $\Delta(\mu_0)$ curve is linear, so $\beta^* = \beta = 0$.

Let $f(t) = g''(t)/(g'(t))^2$, then

$$\frac{\partial^2 \Delta}{\partial \mu_0^2} = \frac{\{g'(\mu_0)\}^2 \{g'(\mu_1)\}^2 \{f(\mu_0) - f(\mu_1)\}}{\{g'(\mu_1)\}^3} .$$

Link function g is an increasing function, so the denominator and $\{g'(\mu_0)\}^2\{g'(\mu_1)\}^2$ will be positive. Hence, for $\beta > 0$, Δ will be strictly concave, convex or linear if f is increasing, decreasing or constant respectively. Function f will be increasing, decreasing or constant if $F(t)$ is convex, concave or linear, respectively. Where $F(t)$ is

$$F(t) = \int_{-\infty}^{t} f(u)du = \frac{-1}{g'(t)} .$$

For $\beta < 0$, the concavity or convexity of Δ will be reversed. So by Lemma C.6, if $F(t)$ is convex, concave or linear then $|\beta^*| < |\beta|$, $|\beta^*| > |\beta|$ or $\beta^* = \beta$ respectively. □

Corollary C.8.
The adjusted and unadjusted parameters of treatment effect coincide ($\beta = \beta^*$) if g is a linear or log link. If g is the logistic link then $\beta \geq \beta^*$.

Proof. If $g(t) = at + b$ then $g'(t) = a$ and so $F(t) = -1/a$, then by Theorem C.7 $\beta = \beta^*$. If $g(t) = c\log(at + b) + d$ then $g'(t) = \frac{ac}{at+b}$ and so $F(t) = \frac{-1}{c}t + \frac{-b}{ac}$, then by Theorem C.7 $\beta = \beta^*$.

If $g(t) = \log\left(\frac{t}{1-t}\right)$ then $F(t) = t^2 - t$ and $\frac{\partial^2 F}{\partial t^2} = 2 > 0$. Then by Theorem C.7 $|\beta^*| < |\beta|$. □

Neuhaus & Jewell [7] also give an approximate expression for the unadjusted treatment effect parameter in terms of the adjusted treatment effect parameter, for adjusted treatment effect parameter close to zero. This is given in Theorem C.9

Theorem C.9.

For β close to zero, β^* is approximately independent of X and

$$\beta^* \approx \beta H'(\beta) = \beta g'(E[g^{-1}(\psi)])E[1/g'(g^{-1}(\psi))]$$

where $H(\beta)$ is the function of β that gives β^*: $H(\beta) = g(\mu_1^*) - g(\mu_0^*) = \beta^*$.

Proof. This is given by the first-order Taylor expansion of $\beta^* = g(\mu_1^*) - g(\mu_0^*) = H(\beta)$. □

C.3 Approach of Robinson et al. for precision with binary outcomes and covariates

In this section I present details of the work of Robinson et al. [5] on precision with binary outcomes and covariates.

Let Y be a binary outcome variable, X a binary variable identifying treatment arm and Z a binary covariate. By randomisation, X and Z are independent. Consider the GLM defined by

$$E(Y|X,Z) = g^{-1}(\eta) = h(\eta) \ ,$$
$$\eta = \alpha + \beta X + \gamma Z$$

and Y follows an exponential distribution, so that β is the treatment effect adjusted for the covariate Z. Also consider an alternative GLM with linear predictor

$$\eta = \alpha^* + \beta^* X$$

so that β^* is the unadjusted treatment effect.

Robinson et al. [5] first define several conditional probabilities as follows:

$$\begin{aligned} p_i &= pr(Y = 1|X = i) \text{ for } i = 0, 1 \\ p_{ij} &= pr(Y = 1|X = i, Z = j) \text{ for } i, j = 0, 1 \\ c_{ij} &= pr(Z = j|X = i) \text{ for } i, j = 0, 1 \end{aligned} \tag{C.8}$$

By usual probability rules we then have the following relationships:

$$p_i = c_{i0}p_{i0} + c_{i1}p_{i1} \qquad \text{and} \qquad c_{i0} + c_{i1} = 1 \text{ for } i = 0, 1$$

X and Z are independent by randomisation, so we also have:

$$c_{0j} = c_{1j} \qquad \text{for } j = 0, 1$$

So let $c = c_{01} = c_{11} = pr(Z = 1)$ and $(1 - c) = c_{00} = c_{10} = pr(Z = 0)$.

Given the GLMs defined above, we can express the treatment effect parameters in terms of conditional probabilities via the link function as follows:

$$\begin{aligned}\beta &= g(p_{10}) - g(p_{00}) = g(p_{11}) - g(p_{01}) \\ \beta^* &= g(p_1) - g(p_0)\end{aligned}$$

Also note that

$$p_{10} = g^{-1}(\alpha + \beta) \qquad \text{and} \qquad p_{11} = g^{-1}(\alpha + \beta + \gamma)\ .$$

The asymptotic relative precision (ARP) of the adjusted estimate of treatment effect to the unadjusted estimate of treatment effect is given by

$$ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{var(\hat{\beta}|X)^{-1}}{var(\hat{\beta}^*|X)^{-1}} = \frac{var(\hat{\beta}^*|X)}{var(\hat{\beta}|X)}\ .$$

By standard likelihood theory results, we can express the asymptotic variances of the unadjusted and adjusted treatment effect estimators in terms of the sample size in each treatment arm, the conditional probabilities given in equations C.8 and the link function:

$$\begin{aligned}var(\hat{\beta}|X) = &\left[\left\{\frac{1}{N_0 c_{00} k(p_{00})} + \frac{1}{N_1 c_{10} k(p_{10})}\right\}^{-1}\right. \\ &\left. + \left\{\frac{1}{N_0 c_{01} k(p_{01})} + \frac{1}{N_1 c_{11} k(p_{11})}\right\}^{-1}\right]^{-1} \qquad \text{(C.9)} \\ var(\hat{\beta}^*|X) = &\frac{1}{N_0 k(p_0)} + \frac{1}{N_1 k(p_1)}\end{aligned}$$

where $k(p) = [p(1-p)\{g'(p)\}^2]^{-1}$ and N_0 and N_1 are the sample sizes in each treatment arm.

Robinson et al. [5] then show that for a linear or log link function the ARP of the adjusted to unadjusted treatment effect estimators is greater than one. So when using a GLM for a binary outcome and covariate with a linear or log link function, the adjusted estimator always has smaller asymptotic variance than the unadjusted estimator. The details of the method are given in Theorem C.10. The result is achieved separately for the two forms of link function, by defining an appropriate random variable and function of that variable and then using Jensen's inequality.

Theorem C.10.
If the link function g is linear ($g(p) = p$) or log ($g(p) = \log(p)$) then $ARP(\hat{\beta}$ to $\hat{\beta}^*) \geq 1$.

Proof. If $g(p) = p$ then $k(p) = [p(1-p)]^{-1}$ and we have

$$\begin{aligned} var(\hat{\beta}|X) = &\left[\left\{ \frac{p_{00}(1-p_{00})}{N_0 c_{00}} + \frac{p_{10}(1-p_{10})}{N_1 c_{10}} \right\}^{-1} \right. \\ &\left. + \left\{ \frac{p_{01}(1-p_{01})}{N_0 c_{01}} + \frac{p_{11}(1-p_{11})}{N_1 c_{11}} \right\}^{-1} \right]^{-1} \\ var(\hat{\beta}^*|X) = &\frac{p_0(1-p_0)}{N_0} + \frac{p_1(1-p_1)}{N_1} \end{aligned}$$

Define

$$f(W) = \left[\frac{W(1-W)}{N_0} + \frac{(W+\beta)[1-(W+\beta)]}{N_1} \right]^{-1}$$

with W a random variable taking value p_{01} with probability $c = c_{01} = c_{11}$ and value p_{00} with probability $(1-c) = c_{00} = c_{10}$, so

$$E(W) = cp_{01} + (1-c)p_{00} = p_0 \ .$$

Then we have

$$\begin{aligned} & f(E(W)) = f(p_0) = [var(\hat{\beta}^*|X)]^{-1} \\ \text{and} \qquad & E(f(W)) = cf(p_{01}) + (1-c)f(p_{00}) = [var(\hat{\beta}|X)]^{-1} \ . \end{aligned}$$

And so

$$ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{E(f(W))}{f(E(W))} \ .$$

Let $f(W) = [m(W)]^{-1}$. Then $m(W)$ is the sum of two concave functions and so it is concave. Therefore $f(W)$ is a convex function so by Jensen's inequality $f(E(W)) \leq E(f(W))$ and

$$ARP(\hat{\beta} \text{ to } \hat{\beta}^*) \geq 1 \ .$$

If $g(p) = \log(p)$ then $k(p) = p/(1-p)$. Now define

$$f(W) = \left[\frac{1-W}{N_0 W} + \frac{1-rW}{N_1 rW} \right]^{-1}$$

with $r = \exp(\beta)$ and W a random variable taking value p_{01} with probability $c = c_{01} = c_{11}$ and value p_{00} with probability $(1-c) = c_{00} = c_{10}$. Then again we have

$$\begin{aligned} & f(E(W)) = f(p_0) = [var(\hat{\beta}^*|X)]^{-1} \\ & E(f(W)) = cf(p_{01}) + (1-c)f(p_{00}) = [var(\hat{\beta}|X)]^{-1} \\ \text{and} \qquad & ARP(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{E(f(W))}{f(E(W))} \ . \end{aligned}$$

Let $f(W) = [m(W)]^{-1}$. Then $m(W)$ is the sum of two concave functions and so it is concave. Therefore $f(W)$ is a convex function so by Jensen's inequality $f(E(W)) \leq E(f(W))$ and

$$ARP(\hat{\beta} \text{ to } \hat{\beta}^*) \geq 1 \ .$$

□

As in Theorem C.11, Robinson et al. [5] find a condition for the ARP to be less than one. The method is similar to that used in the proof of Theorem C.10; by defining a particular random variable and function, and using Jensen's and Minkowski's inequalities. Finally, as in Corollary C.12, this condition is shown to apply for logit, probit, complementary log-log and generalised logistic link functions. So when using a GLM for a binary outcome and covariate with these link functions, the adjusted estimator always has greater asymptotic variance than the unadjusted estimator.

Theorem C.11.
If the $k(p) = [p(1-p)\{g'(p)\}^2]^{-1}$ *is a concave function then* $ARP(\hat{\beta}$ *to* $\hat{\beta}^*) \leq 1$.

Proof. Let $a_{ij} = M_i c_{ij} k(p_{ij})$, then

$$var(\hat{\beta}|X) = \left\{ \left(\frac{1}{a_{00}} + \frac{1}{a_{10}} \right)^{-1} + \left(\frac{1}{a_{01}} + \frac{1}{a_{11}} \right)^{-1} \right\}^{-1} .$$

Let W be a random variable taking value p_{01} with probability $c = c_{01} = c_{11}$ and value p_{00} with probability $(1-c) = c_{00} = c_{10}$, so

$$E(W) = cp_{01} + (1-c)p_{00} = p_0 .$$

Then we have

$$k(E(W)) = k(p_0)$$
$$\text{and } E(k(W)) = ck(p_{01}) + (1-c)k(p_{00}) .$$

If $k(W)$ is concave then $k(E(W)) \geq E(k(W))$ by Jensen's inequality and so

$$N_0 k(p_0) \geq N_0(c_{00}k(p_{00}) + c_{01}k(p_{01})) = a_{00} + a_{01} .$$

Similarly we can obtain

$$N_1 k(p_1) \geq N_1(c_{10}k(p_{10}) + c_{11}k(p_{11})) = a_{10} + a_{11} .$$

Therefore

$$var(\hat{\beta}^*|X) = \frac{1}{N_0 k(p_0)} + \frac{1}{N_1 k(p_1)} \leq \frac{1}{a_{00} + a_{01}} + \frac{1}{a_{10} + a_{11}}$$
$$\Rightarrow \qquad \left[var(\hat{\beta}^*|X) \right]^{-1} \geq \left[\frac{1}{a_{00} + a_{01}} + \frac{1}{a_{10} + a_{11}} \right]^{-1} .$$

By Minkowski's inequality

$$\left(\frac{1}{a_{00} + a_{01}} \right)^{-1} + \left(\frac{1}{a_{01} + a_{11}} \right)^{-1} \leq \left(\frac{1}{a_{00} + a_{01}} + \frac{1}{a_{10} + a_{11}} \right)^{-1}$$
$$\Rightarrow \qquad \left[var(\hat{\beta}|X) \right]^{-1} \leq \left[var(\hat{\beta}^*|X) \right]^{-1}$$
$$\Rightarrow \qquad ARP(\hat{\beta} \text{ to } \hat{\beta}^*) \leq 1 .$$

□

Corollary C.12.
If the link function g is the logit, probit, log-log, complementary log-log or generalised logistic (with $0 < \theta \leq 1$) function then $ARP(\hat{\beta}$ to $\hat{\beta}^*) \leq 1$.

Proof. If $g(p) = \log(p/(1-p))$ then $g'(p) = [p(1-p)]^{-1}$ and so $k(p) = p(1-p)$ which is concave. Then, by Theorem C.11 $ARP(\hat{\beta}$ to $\hat{\beta}^*) \leq 1$.

It can be shown [5] that for probit, log-log, complementary log-log and generalised logistic (with $0 < \theta \leq 1$) link functions, $k(p)$ is also concave and so by Theorem C.11 $ARP(\hat{\beta}$ to $\hat{\beta}^*) \leq 1$. □

C.4 Approach of Robinson et al. for power with binary outcomes and covariates

In this section I present details of the work of Robinson et al. [5] on power with binary outcomes and covariates.

Let the GLMs and conditional probabilities be as defined in Section C.3, so that we have unadjusted treatment effect parameter β^* and adjusted treatment effect parameter β. Recall the following probabilities defined in equations C.8:

$$\begin{aligned} p_i &= pr(Y = 1|X = i) \text{ for } i = 0, 1 \\ p_{ij} &= pr(Y = 1|X = i, Z = j) \text{ for } i, j = 0, 1 \\ c_{ij} &= pr(Z = j|X = i) \text{ for } i, j = 0, 1 \end{aligned}$$

Then, as before, by usual probability rules we have the following relationships:

$$p_i = c_{i0}p_{i0} + c_{i1}p_{i1} \qquad \text{and} \qquad c_{i0} + c_{i1} = 1 \text{ for } i = 0, 1$$

X and Z are independent by randomisation, so we also have:

$$c_{0j} = c_{1j} \qquad \text{for } j = 0, 1$$

So again let $c = c_{01} = c_{11} = pr(Z = 1)$ and $(1 - c) = c_{00} = c_{10} = pr(Z = 0)$. Given the GLMs defined, we can again express the treatment effect parameters in terms of conditional probabilities via the link function:

$$\begin{aligned} \beta &= g(p_{10}) - g(p_{00}) = g(p_{11}) - g(p_{01}) \\ \beta^* &= g(p_1) - g(p_0) \end{aligned}$$

Also note that

$$p_{10} = g^{-1}(\alpha + \beta) \qquad \text{and} \qquad p_{11} = g^{-1}(\alpha + \beta + \gamma) \ .$$

We can compare tests of no treatment effect by their asymptotic relative efficiency (ARE). Suppose we have a parameter of interest θ and we wish to test $\theta = \theta_0$. And we have a statistic T that is a consistent estimator of $\mu_T(\theta)$, a monotone function (at least close to $\theta = \theta_0$) of θ. A definition of ARE called the Pitman efficiency of two such statistics T_1 and T_2 that are asymptotically normally distributed with mean $\mu_T(\theta)$ and variance $\sigma_T^2(\theta)$is given by [28]:

$$ARE(T_1 \text{ to } T_2 \text{ at } \theta = \theta_0) = \left[\frac{\mu'_{T_1}(\theta_0)}{\mu'_{T_2}(\theta_0)}\right]^2 \left[\frac{\sigma^2_{T_2}(\theta_0)}{\sigma^2_{T_1}(\theta_0)}\right] \tag{C.10}$$

We can test the null hypothesis $\beta = 0$ by the adjusted or unadjusted Wald test statistics

$$\frac{\hat{\beta}}{\sqrt{var(\hat{\beta})}} \qquad \text{or} \qquad \frac{\hat{\beta}^*}{\sqrt{var(\hat{\beta}^*)}}$$

which have asymptotic normal distributions with means $\beta/\sqrt{var(\hat{\beta})}$ and $\beta^*/\sqrt{var(\hat{\beta}^*)}$ and variances 1. Substituting the adjusted and unadjusted Wald test statistics into the definition of ARE in equation C.10 we obtain the ARE of a test of no treatment effect based on the adjusted estimator to a test based on the unadjusted estimator [5]:

$$ARE(\hat{\beta} \text{ to } \hat{\beta}^*) = \left[\lim_{\beta\to 0}\left(\frac{d}{d\beta}\beta\right)\left(\frac{d}{d\beta}\beta^*\right)^{-1}\right]^2 \left[\lim_{\beta\to 0}\frac{var(\hat{\beta}^*|X)}{var(\hat{\beta}|X)}\right] \tag{C.11}$$

where we are considering β and β^* as functions of β.

The following theorem details the method of Robinson et al. [5] for establishing that the ARE of a test of no treatment effect based on the adjusted estimator to a test based on the adjusted estimator is greater than one, for all link functions, if the covariate is prognostic. If the covariate is independent of outcome (i.e. not prognostic) then the ARE is one. The method of Robinson et al. [5] expresses each of the two factors making up the ARE in terms of the link function and the probabilities defined above. From the Cauchy-Schwarz inequality, an appropriately defined random variable and Jensen's inequality, this expression for the ARE is shown to be greater than or equal to one.

Theorem C.13.
Given the defined GLMs, for any link function $ARE(\hat{\beta} \text{ to } \hat{\beta}^*) \geq 1$.
$ARE(\hat{\beta} \text{ to } \hat{\beta}^*) = 1$ *if and only if* Z *is independent of* Y.

Proof. Firstly considering the first factor of the expression for $ARE(\hat{\beta} \text{ to } \hat{\beta}^*)$ as given in equation C.11:

We have the trivial result $\frac{d}{d\beta}\beta = 1$. Then noting that p_0 is not a function of β, we have

$$\frac{d}{d\beta}\beta^* = \frac{d}{d\beta}g(p_1) = g'(p_1)\frac{d}{d\beta}p_1 \ .$$

Recall that

$$p_1 = (1-c)p_{10} + cp_{11} = (1-c)g^{-1}(\alpha+\beta) + cg^{-1}(\alpha+\beta+\gamma)$$

and so we have

$$\begin{aligned}\frac{d}{d\beta}\beta^* &= g'(p_1)\left(\frac{1-c}{g'(g^{-1}(\alpha+\beta))} + \frac{c}{g'(g^{-1}(\alpha+\beta+\gamma))}\right) \\ &= g'(p_1)\left(\frac{1-c}{g'(p_{10})} + \frac{c}{g'(p_{11})}\right) .\end{aligned}$$

As $\beta \to 0$, $p_{10} \to p_{00}$, $p_{11} \to p_{01}$ and $p_1 \to p_0$, so

$$\lim_{\beta\to 0}\frac{d}{d\beta}\beta^* = g'(p_0)\left(\frac{1-c}{g'(p_{00})} + \frac{c}{g'(p_{01})}\right) .$$

So the first factor in the ARE is

$$\left[\lim_{\beta\to 0}\left(\frac{d}{d\beta}\beta\right)\left(\frac{d}{d\beta}\beta^*\right)^{-1}\right]^2 = \left[g'(p_0)\left(\frac{1-c}{g'(p_{00})} + \frac{c}{g'(p_{01})}\right)\right]^{-2} . \tag{C.12}$$

Now considering the second factor of the expression for $ARE(\hat{\beta} \text{ to } \hat{\beta}^*)$ as given in equation C.11:

Using the asymptotic variance formulae given in equations C.9, and that as $\beta \to 0$, $p_{10} \to p_{00}$, $p_{11} \to p_{01}$ and $p_1 \to p_0$:

$$\lim_{\beta\to 0} var(\hat{\beta}^*|X) = \lim_{\beta\to 0}\left(\frac{1}{N_0 k(p_0)} + \frac{1}{N_1 k(p_1)}\right) = \{k(p_0)\}^{-1}\left(\frac{1}{N_0} + \frac{1}{N_1}\right)$$

$$\text{and} \qquad \lim_{\beta\to 0} var(\hat{\beta}|X) = \{(1-c)k(p_{00}) + ck(p_{01})\}^{-1}\left(\frac{1}{N_1} + \frac{1}{N_0}\right)$$

$$\therefore \qquad \lim_{\beta\to 0}\frac{var(\hat{\beta}^*|X)}{var(\hat{\beta}|X)} = \frac{(1-c)k(p_{00}) + ck(p_{01})}{k(p_0)}$$

Now letting $A = \{g'(p_{00})\}^{-1}$ and $B = \{g'(p_{01})\}^{-1}$, by equations C.12 and we can express the ARE in terms of the link function and defined probabilities:

$$ARE(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{p_0(1-p_0)\left\{\frac{(1-c)A^2}{p_{00}(1-p_{00})} + \frac{cB^2}{p_{01}(1-p_{01})}\right\}}{\{(1-c)A + cB\}^2} \tag{C.13}$$

To find an inequality involving the ARE, first note that it can by shown by the Cauchy-Schwarz inequality:

$$\{(1-c)A+cB\}^2 \leq \left\{\frac{(1-c)A^2}{p_{00}(1-p_{00})} + \frac{cB^2}{p_{01}(1-p_{01})}\right\}\{(1-c)p_{00}(1-p_{00}) + cp_{01}(1-p_{01})\}$$

Now consider a random variable W that takes value p_{01} with probability c and value p_{00} with probability $1-c$, then

$$E(W) = p_0 \ , \ p_0(1-p_0) = E(W)(1-E(W))$$

$$\text{and} \qquad E(W(1-W)) = (1-c)p_{00}(1-p_{00}) + cp_{01}(1-p_{01}) .$$

Since W^2 is a strictly convex function of W, by Jensen's inequality:

$$\begin{aligned} & E(W^2) \geq E(W)^2 \\ \Leftrightarrow \quad & E(W) - E(W)^2 \geq E(W) - E(W^2) \\ \Leftrightarrow \quad & E(W)(1 - E(W)) \geq E(W(1-W)) \\ \Leftrightarrow \quad & p_0(1-p_0) \geq (1-c)p_{00}(1-p_{00}) + cp_{01}(1-p_{01}) \end{aligned}$$

So we have

$$\{(1-c)A + cB\}^2 \leq p_0(1-p_0)\left\{\frac{(1-c)A^2}{p_{00}(1-p_{00})} + \frac{cB^2}{p_{01}(1-p_{01})}\right\}$$

which by equation C.13 gives our result

$$ARE(\hat{\beta} \text{ to } \hat{\beta}^*) \geq 1 \ .$$

Note that $ARE(\hat{\beta} \text{ to } \hat{\beta}^*) = 1$ if and only if the variable W is constant, which is equivalent to $p_{00} = p_{01}$. So $ARE(\hat{\beta} \text{ to } \hat{\beta}^*) = 1$ if and only if Z is independent of Y. □

C.5 Approach of Neuhaus

In this section I present details of the work of Neuhaus [4].

Let Y be an outcome variable, X a binary variable identifying treatment arm (taking value 1 with probability $p = \frac{1}{2}$) and Z a covariate (here, we consider Z as one-dimensional but the methods generalise to Z as a multi-dimensional vector). By randomisation, X and Z are independent. Assume that $E(Z) = \mu_Z$ and $var(Z) = \sigma_Z^2$. Consider the GLM that includes the covariate, defined by

$$\begin{aligned} E(Y|X,Z) &= g^{-1}(\eta) = h(\eta) \ , \\ \eta &= \alpha + \beta X + \gamma Z \end{aligned}$$

and Y follows a distribution in the exponential family. So β is the adjusted treatment effect parameter. Assume the link function g is strictly monotone increasing and differentiable. For convenience, define the *covariate effect* $\psi = \gamma Z$ so $E(\psi) = \gamma\mu_Z$ and $var(\psi) = \gamma^2\sigma_Z^2$.

Also consider a GLM omitting the covariate Z, so we use the same GLM except with linear predictor

$$\eta^* = \alpha^* + \beta^* X$$

so that β^* is the unadjusted treatment effect parameter.

Recall the probability density function of Y is of the form [79]

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

for specific functions $a(.)$, $b(.)$ and $c(.)$. And the mean and variance of Y are

$$E(Y) = \mu = b'(\theta) \qquad \text{and} \qquad var(Y) = b''(\theta)a(\phi) = Va(\phi)\ .$$

The asymptotic relative efficiency (ARE) of a test of no treatment effect based on the unadjusted estimator to a test based on the adjusted estimator is given by

$$ARE(\hat{\beta}^* \text{ to } \hat{\beta}) = \left[\lim_{\beta\to 0}\left(\frac{d}{d\beta}\beta^*\right)\left(\frac{d}{d\beta}\beta\right)^{-1}\right]^2 \left[\lim_{\beta\to 0}\frac{var(\hat{\beta}|X)}{var(\hat{\beta}^*|X)}\right] \tag{C.14}$$

where we are considering β and β^* as functions of β. (Note that this ARE is the inverse of that described in the previous section; unadjusted and adjusted are interchanged. In both sections, the ARE considered is consistent with the source paper.)

Theorem C.14.
The asymptotic relative efficiency (ARE) of a test of no treatment effect based on the unadjusted estimator to a test based on the adjusted estimator is less than or equal to one: $ARE(\hat{\beta}^* \text{ to } \hat{\beta}) \leq 1$.

Proof. Firstly, consider the first factor in the expression for the ARE. Neuhaus & Jewell [7] show that

$$\lim_{\beta\to 0}\frac{\partial\beta^*}{\partial\beta} = g'\left(E\left[g^{-1}(\alpha+\psi)\right]\right)E\left[\frac{1}{g'\left(g^{-1}(\alpha+\psi)\right)}\right] = g'\left(E[\mu_0]\right)E\left[\frac{1}{g'(\mu_0)}\right] \tag{C.15}$$

and we have the trivial result $\frac{d}{d\beta}\beta = 1$. So the first factor of the ARE is given by

$$\left[\lim_{\beta\to 0}\left(\frac{d}{d\beta}\beta^*\right)\left(\frac{d}{d\beta}\beta\right)^{-1}\right]^2 = g'\left(E[\mu_0]\right)^2 E\left[\frac{1}{g'(\mu_0)}\right]^2 . \tag{C.16}$$

To calculate the second factor of the ARE, we need to find the variances of the estimators of the unadjusted and adjusted treatment effect parameters. Firstly, introduce the following notation for the vector of parameters in the unadjusted model:

$$\beta^* = (\alpha^*, \beta^*)$$

And $\hat{\beta}^*$ is the maximum likelihood estimator of β^*. It has been shown by White [80] that β^* is asymptotically normally distributed with covariance matrix given by the matrix product

$$var(\hat{\beta}^*) = \mathbf{A}^{-1}(\hat{\beta}^*)\mathbf{B}(\hat{\beta}^*)\mathbf{A}^{-1}(\hat{\beta}^*) \tag{C.17}$$

where the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\mathbf{A_{ij}}\left(\beta^*\right) = E_{X,Z}\left[E_{Y|X,Z}\left[\frac{\partial^2 \log P_F(Y=y|\beta^*,X)}{\partial\beta^*_{\mathbf{i}}\partial\beta^*_{\mathbf{j}}}\right]\right]$$

and $$\mathbf{B_{ij}}\left(\beta^*\right) = E_{X,Z}\left[E_{Y|X,Z}\left[\frac{\partial \log P_F(Y=y|\beta^*,X)}{\partial\beta^*_{\mathbf{i}}} \times \frac{\partial \log P_F(Y=y|\beta^*,X)}{\partial\beta^*_{\mathbf{j}}}\right]\right] .$$

Expectations in the above expressions are with respect to the adjusted model, while P_F are probabilities defined by the unadjusted model. Calculating the appropriate derivatives, we obtain the following matrices for the vector of parameters of the unadjusted model:

$$\mathbf{A}\left(\beta^*\right) = E_{X,Z}\left[E_{Y|X,Z}\left[\frac{1}{\phi^* V^*[g'(\mu^*)]^2}\mathbf{M}(X)\right]\right] = E_{X,Z}\frac{1}{\phi^* V^*[g'(E\mu_0)]^2}\mathbf{M}(X)$$

$$\mathbf{B}\left(\beta^*\right) = E_{X,Z}\left[E_{Y|X,Z}\left[\frac{(Y-E\mu_0)^2}{[\phi^* V^* g'(E\mu_0)]^2}\mathbf{M}(X)\right]\right] = E_{X,Z}\frac{\phi V + (\mu - E\mu_0)^2}{[\phi^* V^* g'(E\mu_0)]^2}\mathbf{M}(X)$$

where V^* and ϕ^* are the variance function and scale factor under the unadjusted model, V and ϕ are the variance function and scale factor under the adjusted model and the matrix $\mathbf{M}(X)$ is

$$\mathbf{M}(X) = \begin{bmatrix} [c]1 & X \\ X & X^2 \end{bmatrix} .$$

Apply the result of White [80] given in equation C.17 and setting $\beta = 0$ obtains

$$var(\hat{\beta}^*|\beta=0) = \frac{E[\phi V + (\mu_0 - E\mu_0)^2][g'(E\mu_0)]^2}{var(X)} .$$

From McCullagh & Nelder [79] we also obtain the following variance

$$var(\hat{\beta}|\beta=0) = \frac{\phi}{var(X)}\left\{E\left[\frac{1}{V[g'(\mu_0)]^2}\right]\right\}^{-1} .$$

Combining these two results for variance of estimators of treatment effect parameters, we have the following expression for the second factor of the ARE:

$$\frac{var(\hat{\beta}|\beta=0)}{var(\hat{\beta}^*|\beta=0)} = \frac{1}{E[V+\phi^{-1}(\mu_0-E\mu_0)^2]\left(g'(E\mu_0)\right)^2 E\left[\frac{1}{V(g'(\mu_0))^2}\right]} \qquad \text{(C.18)}$$

Now by evaluating the product of equations C.16 and C.18 we obtain the following expression for the ARE:

$$ARE(\hat{\beta} \text{ to } \hat{\beta}^*) = \frac{E\left[\frac{1}{g'(\mu_0)}\right]^2}{E[V+\phi^{-1}(\mu_0-E\mu_0)^2]E\left[\frac{1}{V(g'(\mu_0))^2}\right]}$$

Note that $\phi^{-1}(\mu_0 - E\mu_0)^2 \geq 0$ and so we have the inequality

$$ARE(\hat{\beta} \text{ to } \hat{\beta}^*) \leq \frac{E\left[\frac{1}{g'(\mu_0)}\right]^2}{E[V]E\left[\frac{1}{V(g'(\mu_0))^2}\right]}$$

Now let $T^2 = \frac{1}{V(g'(\mu_0))^2}$ and $W^2 = V$, so $TW = \frac{1}{g'(\mu_0)}$. Then by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
& E[TW]^2 \leq E[T^2]E[W^2] \\
\Rightarrow \qquad & E\left[\frac{1}{g'(\mu_0)}\right]^2 \leq E\left[\frac{1}{V(g'(\mu_0))^2}\right] E[V] \\
\Rightarrow \qquad & ARE(\hat{\beta} \text{ to } \hat{\beta}^*) \leq 1 \; .
\end{aligned}
$$

So the ARE of a test of no treatment effect based on the unadjusted estimator to a test based on the adjusted estimator is less than or equal to one. □

Appendix D

Numerical method

In Chapter 7 I have presented some equations that must be solved for an unknown. Where these cannot be solved algebraically, we may employ a numerical method. In general, we wish to find a numerical solution to

$$f(x) = 0$$

obtaining a value for x to some pre-determined precision. We must be able to calculate $f(x)$ for any given value of x.

I have chosen to use the bisection method. The basic steps of the bisection algorithm, assuming $f(x)$ is an increasing function of x, are:

1. Start with an interval $[a, b]$, such that $f(a) < 0$ and $f(b) > 0$.
2. Calculate $f(c)$, where $c = \frac{a+b}{2}$.
3. If $f(c) = 0$ then $x = c$, and stop.
4. If $f(c) > 0$, then x must lie between a and c, so set $b = c$ and return to step 2. If $f(c) < 0$, then x must lie between c and b, so set $a = c$ and return to step 2.
5. Repeat steps 2 to 4 until $b - a$ is less than a pre-determined value. Then x is in the interval (a, b).

The following is Stata code for a program to find ICC_{Y^*} (given ICC_Y and Π_0) where

$$\text{ICC}_Y = \frac{\Phi_2\left(\Phi^{-1}(\Pi_0), \Phi^{-1}(\Pi_0), \text{ICC}_{Y^*}\right) - \Pi_0^2}{\Pi_0(1 - \Pi_0)} \ .$$

This program starts with the interval $[0, 1]$. The program returns the midpoint of the interval (a, b) if it is the exact solution, or once $b - a < 10^{-7}$. It will also stop after 100 steps, in case of any unforeseen error.

```
program define probiticc, rclass
syntax , ICC(real) PI(real)

local steps=0
local a=0
local b=1
local prec=0.0000001

while `steps' <100 {
        local c = (`a'+`b')/2    // midpoint

        local fc = (binormal(invnormal(`pi'),invnormal(`pi'),`c')-`pi'^2)/(`pi'*(1-`pi'))
        local fa = (binormal(invnormal(`pi'),invnormal(`pi'),`a')-`pi'^2)/(`pi'*(1-`pi'))
        local fb = (binormal(invnormal(`pi'),invnormal(`pi'),`b')-`pi'^2)/(`pi'*(1-`pi'))

        if `fc'==`icc' | (`b'-`a') < `prec' {
                return scalar iccprobit = `c'
                continue, break
        }

        local steps = `steps' + 1
```

```
        if sign('icc'-'fa')==sign('icc'-'fc') {
                local a = 'c'
        }
        else if sign('icc'-'fb')==sign('icc'-'fc') {
                local b = 'c'
        }

        if 'steps'>99 {
                disp "Couldn't find ICCprobit."
        }

}

end
```

Appendix E

Software

StataCorp. 2013. *Stata Statistical Software: Release 13.* College Station, TX: StataCorp LP.

Raudenbush SW et al. 2011. *Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01).* Available from www.wtgrantfoundation.org.

R Core Team (2014). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Appendix F

Review of the use of covariates in cluster randomised trials: paper in press

This appendix contains the Accepted Author Manuscript of the paper in press titled "Adherence to recommendations on the use of covariates in the design, analysis, and reporting of cluster randomised trials: a review of a sample of 300 trial reports" [60].

Adherence to recommendations on the use of covariates in the design, analysis, and reporting of cluster randomised trials: a review of a sample of 300 trial reports

Neil Wright [a] *, Noah Ivers [b,c], Sandra Eldridge [a], Monica Taljaard [d,e], Stephen Bremner [a]

a. Centre for Primary Care and Public Health, Blizard Institute, Queen Mary University of London, London, United Kingdom
b. Family Practice Health Centre and Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, Toronto, Canada
c. Department of Family and Community Medicine, University of Toronto, Toronto, Canada
d. Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
e. Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada

* Corresponding author: Neil Wright

Centre for Primary Care and Public Health, Blizard Institute,

Yvonne Carter Building, 58 Turner Street, London E1 2AB

n.wright@qmul.ac.uk

+44 (0)7963 789 202

Acknowledements

NI is supported by fellowship awards from the Canadian Institutes of Health Research and the Department of Family and Community Medicine at the University of Toronto.

NW's PhD research is funded by the Medical Research Council.

Abstract

Objective

Reviews of the handling of covariates in trials have explicitly excluded cluster randomised trials. In this study, we review the use of covariates in randomisation, the reporting of covariates, and adjusted analyses in cluster randomised trials (CRTs).

Study Design and Setting

We reviewed a random sample of 300 CRTs published between 2000 and 2008 across 150 English language journals.

Results

Fifty-eight per cent of trials used covariates in randomisation. Only 69 (23%) included tables of cluster and individual level covariates. 55% reported significance tests of baseline balance. Of 207 trials that reported baseline measures of the primary outcome, 155 (75%) subsequently adjusted for these in analyses. Of 174 trials that used covariates in randomisation, 30 (17%) included an analysis adjusting for all those covariates. Of 219 trial reports that included an adjusted analysis of the primary outcome, only 71 (32%) reported that covariates were chosen *a priori*.

Conclusions

There are some marked discrepancies between practice and guidance on the use of covariates in the design, analysis, and reporting of CRTs. It is essential that researchers follow guidelines on the use and reporting of covariates in CRTs, promoting the validity of trial conclusions and quality of trial reports.

Key words: cluster randomised trials; covariates; restricted randomisation; CONSORT statement; adjusted analyses

What is new?

Key findings

- Restricted randomisation was used in around half of cluster randomised trials.
- Few trial reports included tables summarising both cluster level and individual level baseline covariates, and over half of trials reported a significance test of baseline balance.
- Adjusted analyses were common in cluster randomised trials, but authors often did not report whether covariates have been chosen before or after seeing the data.

What this adds to what is known?

- There are some marked discrepancies between practice and guidance in the use of covariates in cluster randomised trials.

What is the implication and what should change now?

- Researchers and editors should note existing guidelines on the use and reporting of covariates in cluster randomised trials.

1. Introduction

In cluster randomised trials (CRTs), pre-existing groups (clusters) of individuals (for example, family practices, hospitals, communities, or schools) are randomised to intervention or control treatment arms. This is in contrast to individually randomised trials, where independent individuals are randomly allocated to treatment arms. A cluster randomised design may be used for one or more of several reasons, including logistical or administrative convenience; to avoid contamination; to improve compliance with treatment; to enable the use of routine data only available at the cluster level; or to evaluate an intervention that is applied at the cluster level. Covariates are commonly used in the design and analysis of randomised trials to promote balance between treatment arms and to improve precision and power [1, 2]. Unlike in individually randomised trials, covariates in CRTs can exist at two levels: that of the cluster and that of the individual. A cluster level covariate describes an intrinsic characteristic of the cluster, and is fixed for all individuals in the same cluster. An individual level covariate describes something about an individual, and may vary across individuals in the same cluster.

Specific recommendations are available with respect to the use of covariates in the design and analysis of randomised controlled trials. For example, the CONSORT guidelines [3] include recommendations for the reporting of covariates, and there is published guidance on adjusting for covariates in analyses [1, 4]. Austin et al. [5] and others [6, 7] have reviewed adherence to recommendations with respect to the handling of covariates in randomised trials, but reviews have excluded CRTs. A review of CRTs in primary care [8] included the use of matching or stratification, and reporting of baseline covariates. In this review we extend the work of Austin et al. [5] to CRTs, where the multi-level nature of CRTs introduces additional complexity in the handling of covariates.

CRTs usually randomise a smaller number of units, so there is a greater risk of imbalance of covariates between treatment arms. The use of restricted randomisation, via stratification, matching, or minimisation, is therefore recommended to help balance both cluster level covariates and individual level covariates summarised at the cluster level, when available, between treatment arms [9]. Failure to do so may be a missed opportunity to improve validity and power by promoting balance in prognostic covariates [2, 9].

The CONSORT extension for CRTs [10] recommends reporting both cluster level and individual level covariates as applicable in tables comparing baseline characteristics between arms. Failure to report baseline distributions of covariates limits the extent to which a reader of a trial report can assess the validity of a trial's findings and external relevance. There are clear recommendations against significance testing of balance of covariates between treatment arms [1, 11].

Including covariates that are strongly correlated with outcome in an analysis can improve analytic power and statistical precision [1, 12, 13]. A baseline measure of the outcome variable is often strongly correlated with outcome, and should be included in an adjusted analysis because it represents an opportunity to improve the statistical power of the analysis [14]. Covariates used in randomisation should also be adjusted for in an analysis [1, 4].

The effects of adjusting for covariates in CRTs are more complex than in individually randomised trials, as adjusting for covariates may affect both the cluster level and individual level residual variance of outcome. Only limited guidance for choosing covariates in the analysis of CRTs has been published [2, 14]. As for individually randomised trials, choosing covariates *post hoc* can invalidate conclusions, while failure to report the method and justification of choosing covariates prevents readers from assessing the validity of adjusted analyses [4, 11].

In this paper, we review a random sample of published CRTs with respect to the use of covariates in randomisation, the reporting of covariates by treatment arm, and adjustment for covariates in analyses of the primary outcome. In particular, we assess whether: a) covariates were used in randomisation; b) tables of baseline cluster level and individual level covariates were presented; c) significance tests for baseline balance of covariates were avoided; d) an adjusted analysis of the primary outcome was reported; e) an analysis adjusting for baseline measure of the outcome (if available) was reported; f) an analysis adjusting for all covariates used in randomisation was reported; and g) all covariates were reported to have been chosen *a priori*.

2. Methods

2.1 Trial sample

This review used a random sample of 300 CRTs previously identified using a published electronic search strategy [15] implemented in Medline. This sample has been used to assess the impact of the CONSORT extension [10] on the reporting and methodological quality of CRTs [16]. The sample includes CRTs for which the main report was published between 2000 and 2008, across 150 English language journals. The journals include general medical journals (for example: the *New England Journal of Medicine (NEJM)*; the *Journal of the American Medical Association (JAMA)*; *The Lancet*; and the *British Medical Journal (BMJ)*), and various specialty journals (for example: the *British Journal of Psychiatry*; *Diabetes Care*; the *International Journal of Cancer*; and the *Journal of Nutrition*).

2.2 Data abstraction

Two reviewers (NW and NI) independently read and abstracted a defined list of items from each paper. Any discrepancies were resolved by consensus. The review method was piloted on fifteen papers, leading to clarifications to the abstraction strategy. Twenty five items were extracted from each trial report, relating to: sample size and primary outcome type (3 items); use of covariates in randomisation (4 items); reporting covariates at baseline (6 items); adjusting for covariates in an analysis of the primary outcome (8 items); and selection of covariates (2 items). Ten items related directly to recommendations made for handling covariates; others were abstracted to provide general background characteristics of the trial and information about the use of covariates. Definitions of variables abstracted and justification are described below.

2.3 Data abstracted

We describe each covariate as either a cluster or individual level variable. The level is that at which the covariate is most naturally measured. This may not always be the level at which it is reported or used in analysis. For example, in a trial where general practitioners are the clusters and their patients are the individuals, the number of years since registration as a general practitioner is a cluster level covariate, while the age of a patient is an individual level covariate. A covariate measured entirely or mostly on individuals not taking part in the trial, and used to describe some aspect of the cluster, would be counted as a cluster level covariate. For example, in a trial where the cluster is a general practice and only some patients in the practice are recruited to the trial, the average age of all patients in a general practice is a cluster level covariate. If a single covariate was summarised and reported in multiple ways (for example, mean age and proportion of patients above age 65), it was counted as one covariate.

We limited abstraction to analyses of the primary outcome for each trial. The primary outcome was defined to be the first primary outcome identified by the author. If a primary outcome was not identified by the author, then it was identified as: the outcome used in the sample size calculation; the outcome identified by

the paper's title; or the first reported outcome in the paper. To be included in this review, an analysis had to include an estimate of treatment effect plus a measure of uncertainty (such as a confidence interval or standard error), or a P-value of a significance test of no treatment effect.

We characterised covariates used in analyses as one of the following: baseline measure of primary outcome; covariate used in randomisation (for example, stratification, minimization, or matching factors); and any other covariate. When a study reported change from baseline as the primary outcome, we counted this as an attempt to adjust for the baseline measure of outcome, acknowledging that this is not the recommended approach [1]. We considered covariates used in randomisation separately from other covariates because these should be adjusted for in the analysis [1, 4], and to enable comparison with previous methodological reviews [5].

When considering analyses adjusted for covariates other than the baseline measure of outcome or covariates used in randomisation, only the analysis identified by the author as the main adjusted analysis (or the first reported if a main adjusted analysis was not identified) was evaluated. Subgroup analyses were not considered. Any reason given for selecting covariates for multivariable analysis was considered as justification given.

The reported method for selecting covariates for adjusted analyses was considered *post hoc* if the paper described a method where covariates were chosen after observation of any trial data. For example, selecting covariates due to an observed imbalance between treatment groups, or using a covariate selection algorithm, were *post hoc* methods. If the paper reported that covariates were chosen before any data were observed, for example they were pre-specified in an analysis plan, then the reported method of choosing covariates is *a priori*. In all other cases, the method was considered unclear.

3. Analysis

Results are presented using descriptive statistics, and include two sets of results by subgroups which were chosen post hoc for further investigation. We explored whether the use of covariates in randomisation and adjusting for covariates at cluster and individual level varied according to the number of clusters in the trial. We explored whether reporting covariates in baseline tables, using statistical tests of baseline balance, and reporting choosing covariates *a priori* differed for trials published in high impact journals compared to other journals.

In the first stratified analysis, three subgroups were defined by mean number of clusters per treatment arm. Subgroups were formed by the lower quartile, upper quartile, and remaining trials: 5 or fewer clusters per arm; from 5.5 to 23 clusters per arm; and more than 23 clusters per arm. These subgroups contained 75, 147 and 71 trials, respectively. In this analysis, seven trials were excluded as the mean number of clusters per treatment arm could not be calculated, because either the total number of individuals or the number of clusters in the trial was not reported.

In the second stratified analysis, two subgroups were defined by the impact factor of the publishing journal. Impact factors were obtained from journal citation reports (ISI Web of Science, 2009). We defined high impact journals as those with impact factor greater than ten, which in this sample are: the *New England Journal of Medicine (NEJM)*; the *Journal of the American Medical Association (JAMA)*; *The Lancet*; *Annals of Internal Medicine*; *PLOS Medicine*; and the *British Medical Journal (BMJ)*. Fifty-one trial reports in the sample were published in these high impact journals. The remaining 249 trial reports were published in other journals.

4. Results

4.1 Trial characteristics

Characteristics of the 300 trial reports used in this review are summarised in Table 2 of Ivers et al. [16].

4.2 Randomisation

Details with respect to the use of covariates in randomisation are presented in Table 1. In 174 (58.0%) of the 300 trials, at least one covariate was used in randomisation. In 130 (43.3%) trials, at least one covariate other than cluster location and cluster size was used. In 118 (39.3%) trials, only cluster level covariates were used.

In the 124 trials that used cluster level covariates in randomisation, 74 (59.7%) used one cluster level covariate, 24 (19.4%) used two, and 26 (21.0%) used between three and nine. Only 12 trials used individual level covariates in randomisation of which nine (75%) used only one individual level covariate.

Of 277 trials that were conducted in more than one location, authors of 50 (18.1%) attempted to balance for cluster location in randomisation (three out of 277 were unclear). Of 285 trials with varying cluster sizes, authors of 51 (17.9%) attempted to balance cluster size in randomisation (four out of 285 were unclear). Of 263 trials with multiple location and variable cluster size, authors of 12 (4.6%) attempted to balance both location and cluster size in randomisation.

4.3 Reporting of baseline covariates

Of 300 trial reports, 69 (23.0%) included tables reporting both cluster and individual level covariates by treatment arm, while 158 (52.7%) included a table of individual level covariates only. Further details are given in Table 1. The number of cluster level covariates described in a table or in the text ranged from 1 to 19 (if non-zero), with a median of 3. The number of individual level covariates described in a table or in the text ranged from 1 to 28 (if non-zero), with a median of 7.

Out of 300 trial reports, 207 (69%) reported a baseline measure of the primary outcome. A significance test of the balance of a covariate between treatment arms was reported or referred to in 166 (55.3%) trial reports.

4.4 Adjusted analyses

Out of 300 trial reports, 219 (73.0%) included at least one adjusted analysis of the primary outcome. Of 207 trial reports that reported a baseline measure of the primary outcome, 155 (74.9%) included an analysis adjusting for a baseline measure of the outcome. Of 174 trials that used covariates (including cluster size and location) in randomisation, 30 (17.2%) included an analysis adjusting for all covariates used in randomisation. These results are summarised in Table 1. Seven trial reports did not include any unadjusted or adjusted analysis of the primary outcome, according to our criteria.

There were 140 (46.7%) trials that included an analysis adjusting for covariates other than baseline measure of outcome or covariates used in randomisation. Of these 140 adjusted analyses, 7 (5.0%) included only cluster level covariates, while 100 (71.4%) adjusted for only individual level covariates. In 29 (20.7%) analyses both cluster and individual level covariates were included. In four (2.9%) of these trials it was unclear which level of covariates had been included. When included, the number of cluster level covariates ranged from 1 to 8, with a median of 1. Likewise, the number of individual level covariates included in an adjusted analysis ranged from 1 to 28, with a median of 3.

4.5 Choosing covariates

Of the 219 (73.0%) trial reports that included an adjusted analysis of the primary outcome, in 71 (32.4%) authors reported choosing covariates *a priori*, but in 93 (42.4%) it was not reported when covariates adjusted for had been chosen. In 73 (33.3%) trial reports, authors gave some justification for the choice of covariates.

4.6 Subgroup results – number of clusters

The results from the subgroup analyses based on trial size are summarised in Table 2. At least one covariate other than cluster location and cluster size was used in randomisation in 22 out of 75 (29%) trials with five or fewer clusters per treatment arm. In trials with more than 23 clusters per treatment arm, 28 out of 71 (39%) balanced on covariates other than cluster location or cluster size. In the remaining 147 trials, 79 (53.7%) used other covariates in randomisation.

Thirty-three (44%) of the 75 smallest trials included an analysis adjusting for covariates other than baseline measure of outcome and covariates used in randomisation. Five (15%) of those adjusted for cluster level covariates. Of the 71 largest trials (more than 23 clusters per arm), 34 (48%) reported an analysis adjusting for other covariates, and 10 (29%) of those analyses used cluster level covariates. Of the 147 remaining trials, 71 (48%) included an analysis adjusting for other covariates, with 21 (30%) using cluster level covariates.

4.7 Subgroup results – journal impact factor

The results described in this section are summarised in Table 3. Reports published in journals with high impact factors showed higher adherence to three guidelines assessed. In particular, tables reporting both cluster and individual level covariates were included in 24 out of 50 (48%) trial reports in high impact journals, compared with 45 out of 250 (18.0%) trial reports in other journals; a significance test of balance of a covariate between treatment arms was reported or referred to in 17 out of 50 (34%) trial reports in high impact journals, compared to 149 out of 250 (59.6%) trial reports in other journals; and 7 (21%) included an analysis adjusting for all covariates used in randomisation compared to 23 (16.3%) in other journals.

Of 30 trial reports in high impact journals that reported a baseline measure of the primary outcome, twenty-one (70%) included an analysis adjusting for a baseline measure of outcome, compared to 134 (75.7%) out of 177 in other journals. In 19 (51.4%) high impact factor journal reports it was unclear when covariates adjusted for in the analysis had been chosen compared to 74 (40.7%) in other journals.

5. Discussion

In CRTs, baseline balance of covariates may be a greater concern than in individually randomised trials [9]. Despite recommendations to use restricted randomization, especially in smaller CRTs [9], over 40% of authors did not do so. This compares to 64% of authors not using restricted randomisation in a review of CRTs in primary care [8]. Furthermore, attempting to use restricted randomisation to achieve balance on individual level covariates was rare, in spite of the fact that balance of both cluster level and individual level covariates is important in a CRT [9].

While 80% of trials included at least one table reporting covariates by treatment group, less than 25% included tables for both cluster and individual level covariates at baseline, as recommended by CONSORT guidelines [10]. The proportion of trial reports including a table of individual level covariates by treatment arm at baseline was slightly higher in high impact journals (84%). This compares to 96.5% and 96% in the

individually randomised trials reviewed by Austin et al. [5] and Saquib et al. [17] respectively. Those reviews included only papers published in selections of journals with high impact factors. This suggests greater adherence in trial reports in high impact journals, rather than a difference in reporting between CRTs and individually randomised trials.

More than half of sampled CRTs (55.3%) reported a significance test of baseline balance, despite clear recommendations against this practice [1, 11]. Although these recommendations strictly apply to individually randomised trials, the argument extends directly to not testing baseline balance of covariates that can only be imbalanced by chance. This practice appears less common in high impact journal reports (34%) and compares to 38.2% and 46% in the reviews of individually randomised trials by Austin et al. [5] and Saquib et al. [17] respectively.

At least one adjusted analysis of the primary outcome was reported in 73% of CRT reports. Adjusted analyses of the primary outcome using the baseline measure of outcome (as recommended by Murray [14]) were conducted in 74.9% of trials which reported baseline values of the outcome. An analysis adjusting for other covariates (not baseline measure of outcome, or covariates used in randomisation) was reported in almost half (46.7%) of trial reports. This compares to 34.2% of individually randomised trials in the review by Austin et al. [5] which included analyses adjusting for covariates other than baseline measure of outcome and covariates used in randomisation.

In one quarter of adjusted analyses some or all covariates were reportedly chosen *post hoc*, defying guidance for choosing all covariates *a priori* [2]. In 93 (42.5%) cases it was not reported when the covariates adjusted for in the analysis had been chosen. This does not allow any assessment of the validity of selecting covariates, but is less common than in the individually randomised trials considered by Austin et al. [5] and Yu et al. [7] (67% and 75.6% of trial reports). Only a third of CRT reports gave any justification for covariate choice, compared to 41.7% and 73.6% of individually randomised trials in the reviews by Assman et al. [6] and Yu et al. [7] respectively.

This review is limited by the age of the included trial reports, ranging from six to fourteen years old. It is plausible that more recently published CRTs could show different characteristics. In addition, we limited abstraction to one primary outcome and one adjusted analysis from each trial report. Therefore, the number of adjusted analyses actually conducted is almost certainly greater than we report. We did not carry out any hypothesis testing or inferential analysis, as we sought to describe practice and had no *a priori* hypotheses. Similarly, our results by subgroup were descriptive as we had no *a priori* hypotheses. Finally, no inferences were made with respect to the appropriateness of methods used for randomisation or adjusting for covariates in any trial as there is uncertainty regarding ideal strategies.

In summary, there are some marked discrepancies between practice and guidance for CRTs with regard to using covariates in randomisation, reporting covariates, testing balance, and methods of choosing covariates. As with the reviews of individually randomised trials, there is inadequate adherence to clear guidance such as not using statistical tests of baseline balance of covariates, and choosing covariates to be used in an adjusted analysis *a priori*. There appears to be better adherence to some recommendations amongst CRTs published in high impact journals, compared to CRTs reported in other journals. Recommendations to use restricted randomisation and to report both cluster level and individual level covariates are not well followed. It is essential that researchers conducting CRTs follow existing guidelines on the use and reporting of covariates to ensure validity of trial conclusions and aid readers in assessing the quality and results of a trial. Readers should be aware of limitations in trials that do not adhere to such

guidelines. Further research is needed into the effects of adjusting for cluster and individual level covariates in the analysis of CRTs, and further guidance is needed for choosing covariates in CRTs.

1. CPMP. Points to Consider on Adjustment for Baseline Covariates (CPMP/EWP/2863/99). EMEA: London, 2003.

2. Raab GM, Butcher I. Balance in cluster randomized trials. Statistics in Medicine. 2001;20(3):351-65.

3. Schulz KF, Altman DG, Moher D, Grp C. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials. 2010;11.

4. Raab GM, Day S, Sales J. How to select covariates to include in the analysis of a clinical trial. Controlled Clinical Trials. 2000;21(4):330-42.

5. Austin PC, Manca A, Zwarenstein M, Juurlink DN, Stanbrook MB. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. Journal of Clinical Epidemiology. 2010;63(2):142-53.

6. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet. 2000;355(9209):1064-9.

7. Yu L-M, Chan A-W, Hopewell S, Deeks JJ, Altman DG. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. Trials. 2010;11.

8. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. Clinical trials. 2004;1(1):80-90.

9. Ivers NM, Halperin IJ, Barnsley J, Grimshaw JM, Shah BR, Tu K, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. Trials. 2012;13.

10. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Grp C. Consort 2010 statement: extension to cluster randomised trials. British Medical Journal. 2012;345.

11. Senn S. Testing for baseline balance in clinical trials. Statistics in Medicine. 1994;13(17):1715-26.

12. Roozenbeek B, Maas AIR, Lingsma HF, Butcher I, Lu J, Marmarou A, et al. Baseline characteristics and statistical power in randomized controlled trials: Selection, prognostic targeting, or covariate adjustment? Critical Care Medicine. 2009;37(10):2683-90.

13. Hernandez AV, Steyerberg EW, Habbema JDF. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. Journal of Clinical Epidemiology. 2004;57(5):454-60.

14. Murray DM. Design and analysis of group-randomized trials. New York: Oxford University Press; 1998.

15. Taljaard M, McGowan J, Grimshaw JM, Brehaut JC, McRae A, Eccles MP, et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. BMC Medical Research Methodology. 2010;10.

16. Ivers NM, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. BMJ. 2011;343.
17. Saquib N, Saquib J, Ioannidis JPA. Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. BMJ. 2013;347.

Table 1: Use of covariates in randomisation, reporting of covariates, and use of covariates in analysis.

	Number of trials / Relevant trials	(%)	
Use of covariates in randomisation			
Covariates used in randomisation	174/300	(58.0%)	(3/300 unclear)
Types of covariates used in randomisation			
Cluster location	50/277	(18.1%)	(3/277 unclear)
Cluster size	51/285	(17.9%)	(4/285 unclear)
Other:			
Only cluster level covariates	118/300	(39.3%)	
Only individual level covariates	6/300	(2.0%)	
Both cluster and individual level covariates	6/300	(2.0%)	
None	167/300	(55.7%)	
Unclear	3/300	(1.0%)	
Reporting of covariates			
Trial report includes a table reporting:			
Cluster level covariates only	13/300	(4.3%)	
Individual level covariates only	158/300	(52.7%)	
Cluster and individual level covariates	69/300	(23.0%)	
None	60/300	(20.0%)	
Baseline measure of primary outcome reported	207/300	(69.0%)	
Significance test of balance of a covariate between treatment arms	166/300	(55.3%)	
Covariates used in analysis			
Trial report includes:			
An unadjusted analysis of the primary outcome	130/300	(43.3%)	
At least one adjusted analysis of the primary outcome	219/300	(73.0%)	
Trial report includes an analysis of the primary outcome adjusting for:			
Baseline measure of outcome	155/207	(74.9%)	
All covariates used in randomisation	30/174	(17.2%)	
Other covariates	140/300	(46.7%)	
When covariates were reported to be chosen:			
A priori	71/219	(32.4%)	
Post hoc	37/219	(16.9%)	
Both a priori and post hoc	18/219	(8.2%)	
Not reported or unclear	93/219	(42.5%)	
Justification given for the selection of covariates	73/219	(33.3%)	

Table 2: Covariates used in randomisation and covariates used in an adjusted analysis, by subgroup of mean number of clusters per treatment arm.

	Mean number of clusters per treatment arm					
	1 to 5		5.5 to 23		23.5 to 3510.5	
Types of covariates used in randomisation (excluding cluster location and cluster size):						
Only cluster level	18/75	(24.0%)	73/147	(49.7%)	26/71	(36.6%)
Only individual level	2/75	(2.7%)	3/147	(2.0%)	1/71	(1.4%)
Both cluster and individual level	2/75	(2.7%)	3/147	(2.0%)	1/71	(1.4%)
None	52/75	(69.3%)	66/147	(44.9%)	43/71	(60.6%)
Unclear	1/75	(1.3%)	2/147	(1.4%)	0/71	(0.0%)
Covariates (other than baseline measure of outcome and covariates used in randomisation) included in an adjusted analysis:						
Cluster only	0/33	(0.0%)	1/71	(1.4%)	6/34	(17.6%)
Individual only	25/33	(75.8%)	50/71	(70.4%)	23/34	(67.6%)
Both cluster and individual level	5/33	(15.2%)	20/71	(28.2%)	4/34	(11.8%)
Unclear	3/33	(9.1%)	0/71	(0.0%)	1/34	(2.9%)

Table 3: Reporting of covariates and use of covariates in analysis, by subgroup defined by Journal Impact Factor.

	High impact journals (Impact factor > 10)		Other journals	
Reporting of covariates				
Trial report includes a table reporting:				
Cluster level covariates only	2/50	(4.0%)	11/250	(4.4%)
Individual level covariates only	18/50	(36.0%)	140/250	(56.0%)
Cluster and individual level covariates	24/50	(48.0%)	45/250	(18.0%)
None	6/50	(12.0%)	54/250	(21.6%)
Significance test of balance of a covariate	17/51	(34.0%)	149/250	(59.6%)
Covariates used in analysis				
Trial report includes an adjusted analysis of the primary outcome	37/50	(74.0%)	182/250	(72.8%)
Trial report includes an analysis of the primary outcome adjusting for:				
Baseline measure of outcome	21/30	(70.0%)	134/177	(75.7%)
All covariate used in randomisation	7/33	(21.2%)	23/141	(16.3%)
Other covariates	22/50	(44.0%)	118/250	(47.2%)
When covariates were reported to be chosen:				
A priori	12/37	(32.4%)	59/182	(32.4%)
Post hoc	5/37	(13.51%)	32/182	(17.6%)
Both a priori and post hoc	1/37	(2.7%)	17/182	(9.3%)
Unclear	19/37	(51.4%)	74/182	(40.7%)
Justification given for the selection of covariates	12/37	(32.4%)	61/182	(33.5%)

Bibliography

[1] Hauck WW, Neuhaus JM, Kalbfleisch JD, and Anderson S. A consequence of omitted covariates when estimating odds ratios. *Journal of Clinical Epidemiology*, 1991, 44(1):77–81.

[2] Porta M. *A Dictionary of Epidemiology*. Oxford University Press, 5th edition, 2008.

[3] Robinson LD and Jewell NP. Some Surprising Results about Covariate Adjustment in Logistic Regression Models. *International Statistical Review*, 1991, 59(2):227–240.

[4] Neuhaus JM. Estimation efficiency with omitted covariates in generalized linear models. *Journal of the American Statistical Association*, 1998, 93(443):1124–1129.

[5] Robinson LD, Dorroh JR, Lien D, and Tiku ML. The effects of covariate adjustment in generalized linear models. *Communications in Statistics - Theory and Methods*, 1998, 27(7):1653–1675.

[6] Gail MH, Wieand S, and Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 1984, 71(3):431–444.

[7] Neuhaus JM and Jewell NP. A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika*, 1993, 80(4):807–815.

[8] Hauck WW, Anderson S, and Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials*, 1998, 19(3):249–256.

[9] Assmann SF, Pocock SJ, Enos LE, and Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*, 2000, 355(9209):1064–1069.

[10] Austin PC, Manca A, Zwarenstein M, Juurlink DN, and Stanbrook MB. Baseline comparisons in randomized controlled trials. *Journal of Clinical Epidemiology*, 2010, 63(8):940–942.

[11] Yu LM, Chan AW, Hopewell S, Deeks JJ, and Altman DG. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. *Trials*, 2010, 11.

[12] Pocock SJ, Assmann SE, Enos LE, and Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine*, 2002, 21(19):2917–2930.

[13] CPMP. Points to Consider on Adjustment for Baseline Covariates (CPMP/EWP/2863/99). 2003.

[14] Raab GM, Day S, and Sales J. How to select covariates to include in the analysis of a clinical trial. *Controlled Clinical Trials*, 2000, 21(4):330–342.

[15] ICH E Expert Working Group . ICH harmonised tripartite guideline: Statistical principles for clinical trials. *Statistics in Medicine*, 1999, 18(15):1905–1942.

[16] Murray DM. *Design and analysis of group-randomized trials*. Oxford University Press, 1998.

[17] Eldridge S and Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. John Wiley & Sons, 2012.

[18] Kirkwood BR, Hurt L, Amenga-Etego S, Tawiah C, Zandoh C, Danso S, Hurt C, Edmond K, Hill Z, ten Asbroek G, Fenty J, Owusu-Agyei S, Campbell O, Arthur P, and ObaapaVit ATT. Effect of vitamin A supplementation in women of reproductive age on maternal survival in Ghana (ObaapaVitA): a cluster-randomised, placebo-controlled trial. *Lancet*, 2010, 375(9726):1640–1649.

[19] Flannery DJ, Vazsonyi AT, Liau AK, Guo SY, Powell KE, Atha H, Vesterdal W, and Embry D. Initial behavior outcomes for the PeaceBuilders universal school-based violence prevention program. *Developmental Psychology*, 2003, 39(2):292–308.

[20] Eldridge SM, Ukoumunne OC, and Carlin JB. The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions. *International Statistical Review*, December 2009, 77(3):378–394.

[21] Commenges D and Jacqmin H. The intraclass correlation coefficient: distribution-free definition and test. *Biometrics*, 1994, 50(2):517–526.

[22] Donner A and Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold, 2000.

[23] Raab GM and Butcher I. Balance in cluster randomized trials. *Statistics in Medicine*, 2001, 20(3):351–365.

[24] Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, 1997, 2(2):173–185.

[25] Konstantopoulos S. The Impact of Covariates on Statistical Power in Cluster Randomized Designs: Which Level Matters More? *Multivariate Behavioral Research*, 2012, 47(3):392–420.

[26] Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, and Raudenbush S. Optimal Design Plus Empirical Evidence: Documentation for the Optimal Design Software. 2011.

[27] Negassa A and Hanley JA. The effect of omitted covariates on confidence interval and study power in binary outcome analysis: A simulation study. *Contemporary Clinical Trials*, 2007, 28(3):242–248.

[28] Cox DR and Hinkley DV. *Theoretical statistics*. Chapman and Hall, London, 1974.

[29] Hernandez AV, Steyerberg EW, and Habbema JDF. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of Clinical Epidemiology*, 2004, 57(5):454–460.

[30] Lingsma H, Roozenbeek B, and Steyerberg E. Covariate adjustment increases statistical power in randomized controlled trials. *Journal of Clinical Epidemiology*, 2010, 63(12):1391.

[31] Austin PC, Manca A, Zwarenstein M, Juurlink DN, and Stanbrook MB. Covariate adjustment in RCTs results in increased power to detect conditional effects compared with the power to detect unadjusted or marginal effects. *Journal of Clinical Epidemiology*, 2010, 63(12):1392–1393.

[32] Rosenberger WF and Sverdlov O. Handling covariates in the design of clinical trials. *Statistical Science*, 2008, 23(3):404–419.

[33] Kahan BC and Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in medicine*, March 2012, 31(4):328–40.

[34] Gail MH, Tan WY, and Piantadosi S. Tests for no treatment effect in randomized clinical trials. *Biometrika*, 1988, 75(1):57–64.

[35] Kahan BC and Morris TP. Assessing potential sources of clustering in individually randomised trials. *BMC Medical Research Methodology*, 2013, 13(1):58.

[36] Grouin JM, Day S, and Lewis J. Adjustment for baseline covariates: an introductory note. *Statistics in Medicine*, 2004, 23(5):697–699.

[37] Austin PC, Manca A, Zwarenstein M, Juurlink DN, and Stanbrook MB. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. *Journal of Clinical Epidemiology*, 2010, 63(2):142–153.

[38] Senn S. Testing for baseline balance in clinical trials. *Statisics in Medicine*, 1994, 13(17):1715–1726.

[39] Begg CB. Significance tests of covariate imbalance in clinical trials. *Controlled Clinical Trials*, 1990, 11(4):223–225.

[40] Roberts C and Torgerson DJ. Understanding controlled trials - Baseline imbalance in randomised controlled trials. *British Medical Journal*, 1999, 319(7203):185.

[41] Permutt T. Testing for imbalance of covariates in controlled experiments. *Statistics in Medicine*, 1990, 9(12):1455–1462.

[42] Altman DG. Comparability of randomised groups. *The Statistician*, 1985, 34(1):125–136.

[43] Schulz KF, Altman DG, Moher D, and the CONSORT Group . CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials*, 2010, 11(32).

[44] Senn SJ. Base logic: Tests of baseline balance in randomized clinical trials. *Clinical Research and Regulatory Affairs*, 1995, 12(3):171–182.

[45] Frison L and Pocock S. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in Medicine*, 1992.

[46] Bloom HS, Richburg-Hayes L, and Black AR. Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, 2007, 29(1):30–59.

[47] Murray DM and Blitstein JL. Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review*, 2003, 27(1):79–103.

[48] Raudenbush SW, Martinez A, and Spybrook J. Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis*, 2007, 29(1):5–29.

[49] Teerenstra S, Eldridge S, Graff M, de Hoop E, and Borm GF. A simple sample size formula for analysis of covariance in cluster randomized trials. *Statistics in Medicine*, 2012, 31(20):2169–2178.

[50] Nixon RM and Thompson SG. Baseline adjustments for binary data in repeated cross-sectional cluster randomized trials. *Statistics in Medicine*, 2003, 22(17):2673–2692.

[51] Austin PC. A Comparison of the Statistical Power of Different Methods for the Analysis of Repeated Cross-Sectional Cluster Randomization Trials with Binary Outcomes. *International Journal of Biostatistics*, 2010, 6(1).

[52] Neuhaus JM and Kalbfleisch JD. Between-and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, 1998, 54(2):638–645.

[53] Shen L, Shao J, Park S, and Palta M. Between- and within-cluster covariate effects and model misspecification in the analysis of clustered data. *Statistica Sinica*, 2008, 18(2):731–748.

[54] Korendijk EJH, Hox JJ, Moerbeek M, and Maas CJM. Robustness of parameter and standard error estimates against ignoring a contextual effect of a subject-level covariate in cluster-randomized trials. *Behavior Research Methods*, 2011, 43(4):1003–1013.

[55] Klar N and Darlington G. Methods for modelling change in cluster randomization trials. *Statistics in Medicine*, 2004, 23(15):2341–2357.

[56] Cook TD. Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. *Annals of the American Academy of Political and Social Science*, 2005, 599:176–198.

[57] Moerbeek M. Power and money in cluster randomized trials: when is it worth measuring a covariate? *Statistics in Medicine*, 2006, 25(15):2607–2617.

[58] Begg MD and Lagakos S. Loss in efficiency caused by omitting covariates and misspecifying exposure in logistic regression models. *Journal of the American Statistical Association*, 1993, 88(421):166–170.

[59] Hernandez AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, Marmarou A, Choi SC, Lu J, Habbema JDF, and Maas AIR. Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study. *Journal of Neurotrauma*, 2006, 23(9):1295–1303.

[60] Wright N, Ivers N, Eldridge S, Taljaard M, and Bremner S. A review of the use of covariates in cluster randomised trials uncovers marked discrepancies between guidance and practice. *Journal of Clinical Epidemiology*.

[61] Saquib N, Saquib J, and Ioannidis JPA. Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. *British Medical Journal*, 2013, 347.

[62] Eldridge SM, Ashby D, Feder GS, Rudnicka AR, and Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clinical trials*, 2004, 1(1):80–90.

[63] Taljaard M, McGowan J, Grimshaw JM, Brehaut JC, McRae A, Eccles MP, and Donner A. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. *BMC Medical Research Methodology*, 2010, 10(1):15.

[64] Ivers NM, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, Skea Z, Brehaut JC, Boruch RF, Eccles MP, Grimshaw JM, Weijer C, Zwarenstein M, and Donner A. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. *British Medical Journal*, 2011, 343:d5886.

[65] Campbell MK, Piaggio G, Elbourne DR, Altman DG, and Grp C. Consort 2010 statement: extension to cluster randomised trials. *British Medical Journal*, 2012, 345:e5661.

[66] Underwood M, Eldridge S, Lamb S, Potter R, Sheehan B, Slowther AM, Taylor S, Thorogood M, and Weich S. The OPERA trial: protocol for a randomised trial of an exercise intervention for older people in residential and nursing accommodation. *Trials*, 2011, 12.

[67] Underwood M, Lamb SE, Eldridge S, Sheehan B, Slowther AM, Spencer A, Thorogood M, Atherton N, Bremner SA, Devine A, Diaz-Ordaz K, Ellard DR, Potter R, Spanjers K, and Taylor SJC. Exercise for depression in elderly residents of care homes: a cluster-randomised controlled trial. *Lancet*, 2013, 382(9886):41–49.

[68] Priebe S, Burton A, Ashby D, Ashcroft R, Burns T, David A, Eldridge S, Firn M, Knapp M, and McCabe R. Financial incentives to improve adherence to anti-psychotic maintenance medication in non-adherent patients - a cluster randomised controlled trial (FIAT). *BMC Psychiatry*, 2009, 9(61).

[69] Priebe S, Yeeles K, Bremner S, Lauber C, Eldridge S, Ashby D, David AS, O'Connell N, Forrest A, and Burns T. Effectiveness of financial incentives to

improve adherence to maintenance treatment with antipsychotics: cluster randomised controlled trial. *British Medical Journal*, 2013, 347:f5847.

[70] Burton A, Altman DG, Royston P, and Holder RL. The design of simulation studies in medical statistics. *Statistics in Medicine*, 2006, 25(24):4279–4292.

[71] Rodriguez G and Elo I. Intra-class correlation in random-effects models for binary data. *The Stata Journal*, 2003, 3(1):32–46.

[72] Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 1997, 53(3):983–97.

[73] Manor O and Zucker DM. Small sample inference for the fixed effects in the mixed linear model. *Computational Statistics & Data Analysis*, 2004, 46(4):801–817.

[74] Zucker DM, Lieberman O, and Manor O. Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2000, 62(4):827–838.

[75] Westgate PM. On small-sample inference in group randomized trials with binary outcomes and cluster-level covariates. *Biometrical journal*, 2013, 55(5):789–806.

[76] Feiveson AH. Power by simulation. *The Stata Journal*, 2002, 2(2):107–124.

[77] StataCorp . *Stata 13 Base Reference Manual*. Stata Press, 2013.

[78] White IR. simsum: Analyses of simulation studies including Monte Carlo error. *The Stata Journal*, 2010, 10(3):369–385.

[79] McCullagh P and Nelder JA. *Generalized linear models*. Chapman & Hall/CRC, 2000.

[80] White H. Maximum-Likelihood Estimation of Mis-Specified Models. *Econometrica*, 1982, 50(1):1–25.